Interaction mechanisms of intrinsically disordered WH2 repeats with actin by nuclear magnetic resonance spectroscopy

Celia Deville

THÈSE DE DOCTORAT DE
l'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité

Chimie

École doctorale de Chimie Physique et de Chimie Analytique de Paris Centre

Présentée par

Célia DEVILLE

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Etude des états multiples des domaines WH2 en interaction avec l'actine par résonance magnétique nucléaire

à soutenir le 10 juillet 2015

devant le jury composé de :

Mme. Carine VAN HEIJENOORT	Directrice de thèse
M. Bruno KIEFFER	Rapporteur
M. Peter TOMPA	Rapporteur
M. Louis RENAULT	Examinateur
M. Olivier LEQUIN	Examinateur

Contents

List of Figures

List of Tables

Abbreviations

BCA	Bicinchoninic Acid
BSA	Bovine Serum Albumin
CH1/2	Chimera 1/2 of Tβ4 and CibD1
CibD1	1st domain of drosophila Ciboulot
Cobl	Cordon-bleu
DNA	Deoxyribonucleic Acid
DTT	Dithiothréitol
GST	Glutathione S-transferase
IDP	Intrinsically Disordered Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
ITC	Isothermal Titration Calorimetry
LB	Lysogeny Broth
MOI	Multiplicity Of Infection
N-WASP	Neural Wiskott Aldrich Syndrome Protein
N-WASPV2	WH2 bidomain of N-WASP
NMR	Nuclear Magnetic Resonance
NUS	Non-Uniform Sampling
OD	Optical Density
PBS	Phosphate Buffer Saline
RDC	Residual Dipolar Couplings
SDS-PAGE	Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis
Tβ4	Thymosin β4
TCEP	Tris(2-Carboxyethyl)Phosphine
TRX	Thioredoxin
WH2	WASP-Homology 2

Foreword

The actin cytoskeleton is a complex three dimensional network of filaments involved in a wide range of motile processes such as cell motility and adhesion, morphogenesis, synaptic plasticity, cell division and endocytosis. Unlike the rigid bony skeleton, the actin cytoskeleton is highly dynamic and constantly remodeling in response to various stimuli : globular actin proteins self associate in a non covalent manner to form helicoidal filaments which constantly polymerize and depolymerize. These filaments are polarized and preferentially polymerize at the barbed-end and depolymerize at the pointed-end.

In the cell, the assembly of this filament network is tightly regulated by a large number of actin binding proteins. They sequester actin monomers forming a pool of polymerization-ready monomeric actin. They promote assembly and nucleate new filaments in order to, for example, form membrane protrusions for cell migration. They fragment filaments to control their length. They organize super-structures by cross-linking actin filaments. And the list goes on.

Intrinsic disorder plays a major role in actin cytoskeleton remodeling. Various intrinsically disordered protein or proteins containing long disordered fragments were shown to participate in all of the above mentioned functions. Notably, the three known nucleation mechanisms involve largely intrinsically disordered proteins : nucleation promoting factors that activate the Arp2/3 complex, formins and WH2 repeats-based filaments nucleators.

This PhD work focuses on the study of WH2 repeats which are intrinsically disordered versatile regulators of actin assembly. These actin binding modules are present in various actin binding proteins, as single domains or tandem repeats and as sole constituents of the protein or as regions of modular proteins. WH2 domains exert key functions in cellular processes such as morphogenesis, neural development and synaptic plasticity, cell motility and adhesion. In particular, they were shown to be involved in cancer development and metastasis. Some pathogens, such as *Vibrio cholerae*, use WH2 containing proteins to remodel actin cytoskeleton of the host during infection.

WH2 repeats are linked to diverse cellular processes and pathologies and so it is of great interest to decipher the molecular mechanisms by which they remodel the actin cytoskeleton. But the structural characterization of WH2 repeats and of their interactions

with actin is difficult because the WH2:actin complexes are unstable and dynamic objects of high molecular weight and actin spontaneously forms filaments at concentration suitable for structural biology.

All WH2 repeats, despite their highly divergent sequences, fold similarly upon binding actin. A N-terminal amphipatic α-helix interacts with an hydrophobic cleft at actin barbed-end. It is followed by an extended region comprising a consensus LKKT/V motif and a C-terminus of variable length.

Single WH2 domains either sequester actin monomers or promote unidirectional assembly. Previous integrated structural biology studies showed that the function of these single domains is linked to the fuzzyness of the C-terminal region of WH2 repeats in their actin bound forms. The dynamics of this interaction is regulated by scattered small sequence elements distributed on the whole sequence of WH2 domains.

The relationship between sequence, structure/dynamics and function is now well understood for single WH2 repeats but their folding upon binding actin mechanisms remain elusive. A part of this PhD work aims at gaining new insights into these interaction mechanisms. To this end, nuclear magnetic resonance is a technique of choice because it enables investigation of the dynamics and conformational exchange on a large time range. Particular NMR experiments, like relaxation dispersion or chemical exchange saturation transfer even allow detection of "invisible", low-populated, excited or intermediate states.

WH2 tandem repeats display enhanced multifunctionality. In addition to monomer sequestration and promotion of unidirectional actin assembly functions, they nucleate and sever actin filaments, bind filaments barbed-end... The molecular mechanisms responsible for these new functions are still debated.

Actin itself is a highly dynamic protein. NMR analysis of dynamics and conformational changes in actin at atomic resolution, in solution and in presence of various WH2 repeats would be interesting since little is known on the effect of these actin binding proteins on actin dynamics. However, actin cannot be produced in *E. coli*, preventing conventional isotopic labeling in bacteria, and so NMR investigations on actin are so far limited to proton NMR or observation of the binding partner.

The first part of this manuscript presents the biological context of this work. The first chapter briefly introduces intrinsically disordered proteins and their properties. The second chapter presents the actin protein and cytoskeleton and highlights the difficulties of actin sample preparation for *in vitro* experiments. The third chapter describes the WH2 repeats family with particular emphasis on proteins studies in this work.

The second part depicts the experimental approach and methods : production and purification of recombinant WH2 repeats and actin (chapter 4), biochemical characterization of protein stability and protein-protein interactions (chapter 5), NMR experiments to study structure, dynamics and conformational exchange in WH2 repeats (chapter 6).

The third part includes results obtained during this PhD project. Chapter 7 focuses on recombinant actin : characterization of a non-polymerizable actin mutant and isotopic labeling for NMR spectroscopy. The investigation on structural and dynamics properties of free WH2 single domains and bidomains showed diverse behaviors in the family (chapter 8). Finally, exchange between a free and actin-bound single WH2 repeat was studied by chemical exchange saturation transfer experiments. Interaction of WH2 bidomains with actin was followed by NMR and provides insight into cooperativity, domains definition and secondary structure in the bound form (chapter 9).

Acknowledgments :

First of all, I would like to thank all the members of the jury. Bruno Kieffer and Peter Tompa accepted to review this manuscript, I hope they will find it interesting and it will lead to fruitful discussions. Merci à Olivier Lequin d'avoir accepter de présider ce jury et à Louis Renault d'avoir accepté d'être examinateur.

Un immense merci à Carine van Heijenoort, ma directrice de thèse qui m'a accueillie en master puis en thèse, m'a fait découvrir la RMN des biomolécules et m'a donné l'opportunité de travailler sur ce sujet passionnant. J'ai eu l'occasion de faire plein de choses au cours de ces quatre années, tant en biochimie qu'en spectroscopie et j'ai énormément appris sur le travail de recherche et la gestion de projet. Merci pour ton soutien et ta confiance.

Je tiens à remercier tous les membres de l'équipe de RMN de l'ICSN qui ont fait que c'était toujours un plaisir de venir au labo. Merci pour les discussions, scientifiques ou non, pour vos conseils en biochimie, en RMN, en débeuguage de script… Dans le désordre, puisque c'est un thème récurrent de ce manuscript, merci à tous les permanents et gens de passage : Christina, Nelly, Ewen, Nadine, Eric G, Eric J, Annie, Naima, François, jean-Nicolas, Prishila, Sophie, Oriane, Philip, Severine, Safe, Fataneh, JP, Nelson, Anna, Régine, Marion, Arthur, Adrien, Yasmine, Timothee, Abdel. Une pensée toute particulière pour Oriane, Safe, Marion et Adrien, les thésards en plein sprint final.

Merci à Annie Moretto pour son aide précieuse au wetlab les innombrables préparations de milieux de culture, solutions, boites de pétri, gestion des stocks... Merci pour ta gentillesse et ta disponibilité, même pour les préparations de dernière minute !

Merci à Eric Jacquet et Naima Nihri pour les experiences de thermal shift assay sur l'actine recombinante.

Merci à François Bontems pour m'avoir permis de participer au développent de techniques de marquage en cellules d'insecte. Merci aussi pour ta relecture du manuscrit et nombreuses remarques et discussions constructives. Merci également à Annalisa Meola qui a effectué les expériences de marquage en cellules S2.

Merci Alda pour ta bonne humeur, les pauses café, les pasteis et ton appui sans faille pour tout le côté administratif.

Je tiens également à remercier nos collaborateurs de l'équipe de dynamique du cytosquelette de Marie-France Carlier au LEBS. Merci à Louis Renault pour son aide sur les expériences de MST et les nombreuses discussions scientifiques que nous avons pu avoir. Merci à Clotilde Husson pour m'avoir initiée à la production et purification de protéine.

Une partie de ce travail de thèse a été effectué à la plateforme de production de protéine en cellules eucaryotes de l'institut Pasteur dirigée par Stéphane Pêtres. Je tiens à le remercier pour cette collaboration tout au long de la thèse et pour m'avoir fait découvrir la culture de cellules d'insectes. Merci aussi à Christine, Evelyne et Fiona pour leur aide, leurs conseils et leur gentillesse lors de mes passages à Pasteur.

Merci aux copains toujours là pour sortir se changer les idées : Charlotte, les filles du RMN-gang pour les piques-niques en bord de Seine ou à Cité U, Aurélia, les cachanais de région parisienne et les délocalisés, les gros du 13ème. Merci à mes colocataires de Gentilly, sans qui les deux premières années de thèse n'auraient certainement pas eu la même saveur, avec une mention spéciale pour Claire, Fabi et Elo pour les soirées passées à refaire le monde sous les bougainvilliers.

Merci à ma famille, à mes parents pour leurs encouragements à donner le meilleur de moi même et leur soutien indéfectible ("aie confiiiaaance !"). Merci à ma sœur Manon et à mon frère Lucas.

Et pour finir, merci à Damien qui endure au quotidien mon emploi du temps aléatoire de thésarde et mes crises existentielles, pour tout un tas de raisons qui ne regardent que lui.

De tentatives en tâtonnements,
ce n'est qu'en avançant, en
essayant, que l'on pourra
espérer vaincre la dispersion et
l'impuissance.

Mona Chollet
Contre la fascination du désastre

Résumé en français

Contexte biologique

Protéines Intrinsèquement Désordonnées

La biologie structurale est souvent abordée par le prisme du paradigme séquence-structure-fonction : la séquence en acides aminés détermine le repliement tridimensionnel et ce repliement permet d'expliquer la fonction, notamment via des mécanismes de type clé-serrure proposés par Fisher dès la fin du 19ème siècle.

Par ailleurs, de nombreuses protéines ou régions de protéines n'ont pas de structure tridimensionnelle stable et sont pourtant fonctionnelles (Wright and Dyson 1999). Ces protéines intrinsèquement désordonnées ou régions intrinsèquement désordonnées représentent plus d'un tiers du génome des eucaryotes. Il est généralement admis que le désordre est encodé dans la séquence et est lié à une faible fraction de résidus hydrophobes et une forte fraction de résidus hydrophiles, polaires et chargés, ainsi que de glycines et prolines. Le désordre est aussi souvent associé à une moindre complexité de séquence et à des motifs répétés. Il est important de noter que le désordre est relatif, qu'il existe un spectre continu allant des protéines repliées peu dynamiques aux protéines entièrement désordonnées aux propriétés proches des polymères (Habchi et al. 2014).

Les protéines ou régions intrinsèquement désordonnées sont impliquées dans de nombreux processus cellulaires, en particulier dans la signalisation, dans la régulation de la transcription, dans le remodelage du cytosquelette, etc. Les fonctions des IDP/IDRs peuvent être rangées dans plusieurs catégories (van der Lee et al. 2014). Pour certaines IDPs, la fonction est intrinsèquement liée au désordre et ne nécessite pas d'interaction avec d'autres partenaires. C'est le cas des ressorts ou chaines entropiques ainsi que des linkers entre domaines de protéines modulaires. Les autres IDP/IDRs exercent leurs fonctions par le biais d'interactions transitoires ou plus stables. Les régions intrinsèquement désordonnées contiennent souvent des sites de modifications post-traductionnelles accessibles aux enzymes de modification et aux partenaires pour déclencher des cascades de signalisation. Le désordre est particulièrement représenté chez les chaperones aidant le repliement des protéines et des ARN. Des protéines intrinsèquement désordonnées effectrices modulent l'activité de partenaires et peuvent également être impliquées dans des mécanismes d'autoinhibition ou d'autorégulation. Des protéines d'assemblage permettent d'organiser des complexes multiprotéiques. Dans ce cas, le désordre et des conformations étendues

permettent d'avoir une grande surface d'interaction pour un nombre d'acides aminés plus faible que dans le cas de protéines repliées. Certaines IDP lient, séquestrent et neutralisent des petites molécules. Enfin, il a récemment été montré que des IDPs pouvaient être impliquées dans des transitions de phase dans le cytoplasme pour former des organelles sans membrane (Kato et al. 2012).

De nombreuses protéines intrinsèquement désordonnées agissent par le biais d'interactions avec d'autres partenaires, protéines, acides nucléiques, petites molécules ou ions. Souvent, l'interaction s'accompagne d'un repliement partiel ou global de la protéine. Ce repliement peut être modélisé par deux mécanismes extrêmes : la sélection conformationnelle et l'ajustement induit (Wright and Dyson 2009). Dans le premier cas, la protéine intrinsèquement désordonnée est en échange conformationel avec la forme repliée du complexe et cette conformation spécifiquement interagit avec le partenaire. Ce mécanisme permet de minimiser le cout entropique du repliement au cours de l'interaction. Dans le second cas, la protéine interagit d'abord avec le partenaire et se replie à son contact. De nombreuses IDPs forment de manière transitoire certaines des structures secondaires présentes dans la forme liée au partenaire.

Certaines IDP conservent une certaine dynamique lorsqu'elles sont liées à leur partenaire : échange entre deux conformations possibles, région restant désordonnée lors de l'interaction ou même interaction transitoire sans former de complexe stable. On parle alors de complexe flou ("*fuzzy complex*") (Tompa and Fuxreiter 2008). Du fait de leur adaptabilité conformationnelle, certaines IDP peuvent interagir avec plusieurs partenaires, parfois en adoptant des repliements différents.

Le désordre est sur-représenté chez les protéines de régulation du cytosquelette.

L'actine, protéine essentielle du cytosquelette

Les microfilaments d'actine font partie des trois polymères principaux du cytosquelette et forment un réseau tridimensionnel complexe et dynamique au sein de la cellule qui polymérise et dépolymérise constamment. Le cytosquelette d'actine est impliqué dans de nombreuses fonctions physiologiques notamment la morphogénèse, l'établissement et le maintien de la polarité cellulaire, la plasticité neuronale, la motilité et l'adhésion cellulaire, l'endocytose, le trafic intracellulaire, la contraction musculaire (Pollard and Cooper 2009). En raison de ces nombreuses fonctions, le cytosquelette d'actine est impliqué dans des pathologies variées liées à des mutations du monomères d'actine changeant les propriétés du filament, à une mauvaise régulation de l'assemblage des filaments ou à l'utilisation du cytosquelette d'actine par des agents pathogènes externes, microorganismes ou virus, dans leur cycle de développement.

Les eucaryotes possèdent une à six isoformes d'actine exprimées par des gènes différents. La séquence est particulièrement bien conservée avec 86% d'identité de séquence entre l'actine de levure et l'actine α humaine du muscle squelettique. L'actine existe sous

deux formes : l'actine monomérique ou actine G pour globulaire et l'actine F pour filament (Sheterline et al. 1995).

Le monomère d'actine G a une forme de sphère aplatie et est composé de deux domaines principaux eux même composés de deux sous domaines chacun. Un sillon entre les deux domaines abrite un site de liaison à un nucléotide, ATP ou ADP, et à un ion divalent, magnésium en conditions physiologiques et calcium dans la plupart des études structurales (Kabsch et al. 1990). L'extrémité comportant les sous-domaines 1 et 3 est appelée bout barbé, celle comprenant les sous domaines 2 et 4 est appelée bout pointu.

Lorsque la concentration d'actine est au dessus de la concentration critique, l'actine G polymérise pour former des filaments d'actine F. La polymérisation en filaments se fait selon les étapes suivantes *in vitro*. Des ions interagissent de manière spécifique et non spécifique avec les monomères d'actine pour écranter les charges en surface, c'est l'activation. La nucléation consiste en la formation d'oligomères d'actine comprenant à la fois les interactions longitudinales et transverses du filament. C'est l'étape cinétiquement limitante. S'ensuit l'élongation du filament par l'ajout de monomères d'actine aux extrémités du filament, préférentiellement au bout barbé pour lequel la concentration critique est environ cinq fois plus faible qu'au bout pointu. A l'état stationnaire, l'actine ATP polymérise au bout barbé du filament, l'ATP est rapidement hydrolysé en ADP dans le filament, la libération du phosphate est plus lente et finalement, l'actine ADP dépolymérise au bout pointu du filament. Cet état stationnaire est comparé au fonctionnement d'un tapis roulant (*treadmilling*).

Le filament d'actine est un homopolymère non covalent formé par l'assemblage réversible de monomères d'actine. Sa structure hélicoïdale peut être considérée comme une simple hélice gauche avec une rotation de 166°par monomère ou comme une double hélice droite avec une rotation de 28°par monomère. Cette structure est stabilisée à la fois par des interactions longitudinales et transverses entre monomères d'actine. Des données de diffraction de fibres à haute résolution ont permis de montrer que la transition de l'actine G à l'actine F s'accompagne d'une rotation des deux domaines principaux l'un par rapport à l'autre entrainant un aplatissement de la structure de l'actine dans le filament (Oda et al. 2009). Les récents développements en microscopie électronique ont permis d'obtenir des structures du filament d'actine à plus haute résolution et de mieux définir les surfaces en interaction.

Les différentes études structurales portant sur l'actine G et sur l'actine F montrent que l'actine est une protéine particulièrement plastique comportant de nombreuses régions flexible et extrêmement polymorphique, autant sous la forme de monomère que sous la forme de filament (Galkin et al. 2010).

Dans la cellule, l'assemblage des filaments d'actine est régulé de manière très fine dans l'espace et dans le temps par les protéines de liaison à l'actine ou ABP. Ces protéines séquestrent l'actine monomérique, promeuvent l'assemblage, nucléent, fragmentent ou coiffent les filaments... Comprendre ces interactions et leurs spécificités pour différents

états de l'actine est essentiel pour expliquer les nombreux processus cellulaires pilotés par le remodelage du cytosquelette d'actine.

Les études fonctionnelles et structurales nécessitent de préparer des quantités importantes (plusieurs milligrammes) de protéine pure. L'actine représente jusqu'à 20% de la masse de protéines dans les cellules musculaires. Une méthode classique de préparation d'actine est donc l'extraction et la purification à partir de muscle de lapin ou de poulet. Cette méthode permet de purifier de l'actine avec un bon rendement mais est contraignante en terme de logistique, ne permet pas de caractériser des mutants et ne permet d'obtenir que l'isoforme ou le mélange d'isoformes exprimées dans les cellules utilisées, à savoir l'actine α du muscle squelettique pour la purification à partir de muscle de lapin.

Le développement de méthodes pour produire l'actine de manière recombinante sont donc nécessaires pour préparer des échantillons plus variés et adaptés à la question posée. L'actine ne peut pas être exprimée en bactérie en raison de l'absence d'un complexe de protéines chaperones, le complexe CCT, nécessaire pour un repliement correct. La levure ne possède qu'une isoforme d'actine essentielle et la perturbation de son cytosquelette par une actine exogène potentiellement toxique peut compromettre la viabilité des cellules. Il a été montré qu'il était possible de produire de l'actine recombinante en utilisant le système baculovirus/Sf9 avec des rendements suffisants pour des études structurales. Cette méthode a permis de produire différentes isoformes d'actines, des mutants impliqués dans des pathologies comme les myopathies et des mutants non-polymérisables pour faciliter les études structurales.

Les domaines WH2

Les domaines WH2 (Whiskott Aldrick Syndrome Homology 2) et thymosine β (βT) sont une famille de protéines intrinsèquement désordonnées qui interagissent avec l'actine pour réguler sa polymérisation (Renault et al. 2013). Ils sont impliqués dans des processus cellulaires tels que l'embryogenèse, la plasticité neuronale, la motilité cellulaire, l'invasion des cellules cancéreuses et l'infection par certains pathogènes. Ces domaines sont présents de manière isolée ou répétée dans de nombreuses protéines de régulation de la polymérisation de l'actine. Leur longueur varie de 18-20 à 55-60 acides aminés. Il est important de noter que dans le cas de protéines intrinsèquement désordonnées, la délimitation des domaines est plus difficile que dans le cas de domaines structurés.

D'abord présentés comme étant deux familles distinctes, il est désormais clair que les domaines WH2 et thymosine β font partie de la même famille de protéines liant l'actine. Comme pour de nombreuses protéines intrinsèquement désordonnées, les domaines WH2/βT ont des séquences très variables avec seulement 15% de similarités de séquence pour l'ensemble de la famille.

Ils sont principalement caractérisés par un motif consensus L++T/V (ou LKKT/V) où + désigne un résidu chargé positivement. Ce motif est précédé d'une extension N-

terminale qui forme une hélice amphipathique de 2 à 3 tours au contact du sillon hydrophobe entre les sous-domaines 1 et 3 de l'actine. L'extension C-terminale est de longueur variable et dans certains cas, forme une deuxième hélice interagissant avec le point pointu de l'actine.

En l'absence d'actine, les domaines WH2/βT sont dépourvus de structure tridimensionnelle stable. Des analyses RMN ont montré une certaine propension à former l'hélice amphipatique N-terminale pour la thymosine β_4 et pour le premier domaine de Ciboulot en particulier à basse température. Il existe peu d'études caractérisant de manière quantitative les structures résiduelles et la dynamique de domaines WH2/βT en l'absence d'actine.

Les protéines à domaines WH2 peuvent être rangées dans quatre catégories principales. Les domaines thymosine β isolés (comme la thymosine β_4) ou répétés (comme Ciboulot ou la tetrathymosine) interagissent avec l'actine G et séquestrent l'actine monomérique ou promeuvent l'assemblage au bout barbé du filament. Les facteurs de promotion de la nucléation impliqués dans l'activation du complexe Arp2/3 (comme WASP et N-WASP) comportent une extension VCA C-terminale. V désigne le ou les domaines WH2, C est un domaine connecteur impliqué dans l'autoinhibition et la région A est fortement acide. La région CA interagit avec le complexe Arp2/3 qui permet le branchement de filaments d'actine en se liant à un filament mère et la région V composée d'un ou plusieurs domaines WH2 recrute des monomères d'actine pour nucléer le filament fille. Certains domaines WH2 agissent en synergie avec d'autres domaines ou protéines liant l'actine. Enfin, certaines protéines comme Cordon-bleu nucléent l'actine grâce à leurs domaines WH2 répétés.

Pour comprendre la multifonctionnalité de ces domaines, il est nécessaire de caractériser la structure et la dynamique de leurs interactions avec l'actine G ou F. Cela implique d'isoler à des concentration suffisamment hautes, les complexes entre domaines WH2 et un ou plusieurs monomères d'actine G tout en empêchant la polymérisation de l'actine. Plusieurs domaines WH2 ont été cristallisés en complexe avec l'actine grâce à des stratégies variées. Le repliement au contact de l'actine est globalement similaire (Chereau et al. 2005), l'hélice amphipatique interagit avec le sillon hydrophobe entre les sous-domaines 1 et 3 de l'actine, ce qui explique l'inhibition de la polymérisation au bout pointu du filament par ces domaines. La leucine et les résidus chargés positivement du motif LKKT/V interagissent avec des patchs hydrophobes et chargés négativement à la surface de l'actine, la partie C-terminale est plus dynamique et n'est en général pas définie sur les cartes de densité électronique. Des analyses par RMN ont montré que dans le cas de la thymosine β_4 et du premier domaine de Ciboulot, la région C-terminale forme une deuxième hélice au contact de l'actine (Domanski et al. 2004, Hertzog et al. 2004).

Les domaines WH2/βT isolés séquestrent l'actine monomérique (comme la thymosine β_4) ou promeuvent l'assemblage au bout barbé du filament (comme le premier domaine de Ciboulot). Pour mieux comprendre les éléments déterminant la fonction, Didry et al.

(2012) ont caractérisé plusieurs chimères et mutants de la thymosine $\beta 4$ et du premier domaine de Ciboulot par des tests fonctionnels et des méthodes structurales (cristallographie, RMN et SAXS) et ont montré que la fonction de ces domaines était liée à la dynamique de la région C-terminale. Si celle ci est peu dynamique, l'hélice C-terminal coiffe le bout pointu de l'actine et le domaine séquestre l'actine. Si elle est plus dynamique, le bout pointu de l'actine est accessible pour l'ajout au bout barbé d'un filament. La dynamique de la partie C-terminale est déterminée par la nature des interactions polaires et électrostatiques plus ou moins fortes entre les régions centrales et C-terminale des domaines WH2/βT et l'actine. Deux chimères CH1 et CH2 ont des fonctions opposées alors que leurs séquences ne diffèrent que de trois acides aminés dans un linker avant le motif LKKT. La différence de fonction est dans ce cas liée à la possibilité ou non de former un pont salin intermoléculaire avec l'actine. Dans le cas de CH2, le linker contient une lysine qui forme un pont salin intermoléculaire avec un glutamate de l'actine. Cette lysine est remplacée par un glutamate dans CH1.

N-WASP, pour Neural Wiskott Aldrich Syndome Protein, est un facteur de promotion de la nucléation impliqué dans l'activation du complexe Arp2/3. Cette protéine intervient dans de nombreux processus cellulaires comme la motilité cellulaire, la réponse immunitaire, le transport intracellulaire... Contrairement à son homologue WASP exprimé uniquement dans les cellules du système hématopoïétique, N-WASP est ubiquitaire et particulièrement abondamment exprimé dans les cellules nerveuses. N-WASP contient deux domaines WH2 dans son extension VCA C-terminale alors que WASP n'en contient qu'un seul et nuclée de manière plus efficace. Des expériences de titration par fluorescence ont montré que l'actine interagit de manière non coopérative avec les deux domaines WH2 de N-WASP avec deux constantes d'affinité submicromolaires distinctes de 20 et 300nM respectivement. Seul un complexe 1:1 a pu être cristallisé mais la distance entre les deux domaines suggère qu'ils interagissent avec l'actine en formant un dimère longitudinal (Gaucher et al. 2012).

Cordon-bleu, nommé ainsi en raison de son expression concentrée autour de la ligne médiane dans l'embryon de souris, est une protéine conservée chez les vertébrés et impliquée dans formation du tube neural, la morphogénèse neurale, la différenciation gauche/droite, la formation des cellules sensorielles ciliées etc. Cordon-bleu lie l'actine via ses trois domaines WH2 C-terminaux A, B et C. Le premier domaine est précédé d'une région K riche en lysines, et les deuxièmes et troisièmes domaines sont séparés par une région riche en prolines. Husson et al. (2011) ont caractérisé les fonctions de plusieurs fragments montrant ainsi l'importante multifonctionnalité des domaines WH2 de cette protéine : séquestration d'actine ADP, promotion de l'assemblage au bout barbé, nucléation, fragmentation de filaments. La région K est nécessaire pour l'activité de nucléation : KA, KAB et KABC nucléent. La région K et au moins les deux premiers domaines WH2 sont nécessaires pour fragmenter les filaments : KAB et KABC fragmentent.

Il existe peu de données structurales sur l'interaction de ces domaines WH2 avec l'actine. La seule structure a été obtenue grâce à une stratégie de double mutant où deux mutants d'actine non polymérisables forment un dimère non polymérisable interagissant avec le fragment AB de Cordon-bleu. Dans cette structure, les interactions entre les deux monomères d'actine sont très différentes des interactions au sein du filament. La deuxième hélice amphipathique, celle du domaine B, interagit à la fois avec le premier et le deuxième monomère d'actine et l'extension C-terminale du domaine B forme une hélice qui interagit avec le bout pointu de l'actine.

Objectifs de la thèse

Les domaines WH2 jouent un rôle important dans la régulation de l'assemblage de l'actine. La compréhension des mécanismes moléculaires permettant d'expliquer les fonctions variées de ces domaines se heurte à plusieurs obstacles. Tout d'abord, l'actine polymérise et forme des filaments aux concentrations nécessaires pour la cristallogénèse, la RMN ou le SAXS. Il est donc nécessaire de trouver des méthodes pour stabiliser les complexes avec l'actine tout en empêchant la formation des filaments. Les complexes entre domaines WH2 et actine(s) sont hautement flexibles et la cristallographie ne suffit pas à rendre compte des variations de dynamique au sein de différents complexes. La RMN peut apporter des informations sur la dynamique des complexes mais il s'agit d'objets de taille importante pour la RMN, en particulier dans le cas de l'interaction de multidomaines WH2 avec plusieurs actines. Il est donc nécessaire de combiner plusieurs approches pour déchiffrer les mécanismes moléculaires sous-jacents et comprendre comment les domaines WH2 exercent leurs nombreuses fonctions.

Les objectifs de cette thèses se décomposent en trois parties. Une partie concerne la préparation d'actine recombinante en utilisant le système baculovirus/Sf9. Cela permettra de simplifier les protocoles de préparation d'actine pour faciliter les études structurales par rapport à la purification à partir de muscle de lapin. Un autre objectif est de produire un mutant non polymérisable pour les études structurales. Enfin, le développement de méthodes de marquage isotopique des protéines exprimées en cellules d'insecte pourrait permettre d'observer directement l'actine par RMN pour identifier les surfaces d'interactions avec les domaines WH2 en solution et détecter d'éventuelles modifications fines de structure ou de dynamique lors de l'interaction avec les domaines WH2 et de mieux comprendre les états structuraux multiples de l'actine.

Une deuxième partie de ce travail s'attache à caractériser la variabilité structurale et dynamique de plusieurs domaines WH2 isolés et répétés en l'absence d'actine afin de voir s'il est possible de relier ces paramètres à la fonction. On cherche également à mieux comprendre les ensembles conformationnels formés par ces protéines intrinsèquement désordonnées ainsi que leurs modes d'interaction avec l'actine.

Enfin, une dernière partie s'intéresse à la caractérisation de l'interaction avec l'actine.

Dans le cas des domaines WH2 isolés, la relation entre séquence, structure et dynamique du complexe et fonction est maintenant bien comprise. On se concentre donc sur la compréhension du mécanisme de repliement au cours de l'interaction et l'identification de possibles états intermédiaires. Dans le cas de multidomaines WH2, il existe peu de données structurales à haute résolution en solution permettant de caractériser l'interaction avec l'actine. Nous nous concentrons sur l'interaction de deux bidomaines WH2, NWASPV2 et CoblAB, avec l'actine afin de caractériser la structure et la dynamique des complexes.

Approche expérimentale

Préparation des protéines

Quatre domaines WH2 ont été produits de manière recombinante dans *E. coli*, marqués ^{15}N, (^{15}N, ^{13}C) et (^{15}N, ^{13}C, ^{2}H).

Les chimères CH1 et CH2 sont exprimées en fusion avec la GST, purifiées sur des billes de glutathion agarose, la fusion est clivée sur billes par la Prescission et la protéine est purifiée par gel filtration.

Le bidomaine de N-WASP, NWASPV2, est exprimé en fusion avec une étiquette polyhistidine N-terminale. La protéine est purifiée par affinité sur une colonne de nickel, l'étiquette polyhistidine est clivée par la protéase TEV, le tag et la TEV sont capturés sur billes de nickel et la protéine est purifiée par gel filtration.

Le bidomaine AB de Cordon-bleu est exprimé en fusion avec une étiquette polyhistidine et une thiorédoxine N-terminale et avec un streptag C-terminal. La protéine est purifiée par affinité sur une colonne de nickel, la fusion N-terminale est clivée par la Prescission, la fusion et la Prescission sont capturées sur billes de nickel et la protéine est purifiée par gel filtration.

Deux constructions d'actine ont été produites dans le système baculovirus/Sf9 : l'actine 5C sauvage de drosophile et un mutant non polymérisable AP déjà utilisé dans d'autres études structurales pour caractériser l'interaction avec des multidomaines WH2. Une double étiquette Strep a été ajoutée en N-terminal pour une purification par affinité suivie d'une gel filtration.

La production d'actine marquée se fait par transfert direct dans le milieu de marquage après infection des cellules par le baculovirus recombinant.

Caractérisation biochimique

La stabilité des actines recombinantes est évaluée par Thermal Shift Assay, une expérience de dénaturation thermique suivie par fluorescence.

L'affinité de la chimère CH2 pour les actines recombinantes et purifiées à partir de muscle de lapin est déterminée par des expériences de thermophorèse micro-échelle. Les

actines sont marquées à l'aide d'un fluorophore et on suit les variations de diffusion dans un gradient de température en fonction de la quantité de partenaire non marqué.

Spectroscopie RMN

Ce chapitre décrit les méthodes de spectroscopie de résonance magnétique nucléaire utilisées pour la caractérisation des protéines étudiées.

L'attribution des résonances du squelette est effectuée grâce à des expériences de triple résonance hétéronucléaires séquentielles et bidirectionnelles.

La résonance magnétique nucléaire est une technique de choix pour la caractérisation structurale et dynamique des protéines. Les observables en RMN donnent accès à des informations concernant les angles dièdres ce qui permet de déterminer les structures secondaires, les proximités spatiales entre atomes pour déterminer le repliement tertiaire d'une protéine et la dynamique sur une très large gamme de temps allant de la picoseconde à la seconde.

Les protéines intrinsèquement désordonnées ne peuvent pas être décrites par une seule structure tridimensionnelle mais plutôt par un ensemble de conformères flexibles en échange. Les structures secondaires résiduelles et la dynamique de ces ensembles peuvent être importantes pour la fonction de ces IDPs d'où l'importance de les caractériser.

Les déplacements chimiques sont sensibles à l'environnement physico-chimique du noyau considéré lequel dépend de la séquence en acides aminés, des structures secondaires et des conditions expérimentales. Les déplacements chimiques des protéines étudiées ont été analysés pour évaluer les proportions de structures secondaires résiduelles en utilisant deux programmes SSP et δ2d. Les déplacements chimiques expérimentaux sont comparés aux déplacements chimiques attendus dans le cas d'une pelote statistique, d'une hélice α entièrement formée, d'un brin β entièrement formé et d'une polyproline entièrement formée.

Les couplages dipolaires sont également sensibles aux structures secondaires, ainsi qu'aux contacts à longue distance. En solution, ces couplages sont moyennés à zéro en raison de la libre rotation dans le solvant. Il est possible de mesurer des couplages dipolaires résiduels en dissolvant la protéine dans un milieu légèrement anisotrope, ici des mélanges PEG/alcool formant des phases de type cristal liquide qui s'alignent dans le champ magnétique ou des gels de polyacrylamide étirés dans le tube RMN. Les couplages dipolaires résiduels sont analysés à l'aide du logiciel *flexiblemeccano* qui permet de prédire les couplages dipolaires pour des ensembles dont on définit les structures secondaires résiduelles ainsi que leurs populations et les contacts à longue distance.

La vitesse d'échange des protons amides avec le solvant permet d'obtenir des informations sur l'accessibilité des fonctions amides le long de la séquence protéique en comparant les valeurs expérimentales de vitesse d'échange k_{ex} avec les valeurs attendues pour une pelote statistique. Le facteur de protection d'un résidu est le ratio de la vitesse

d'échange des protons amides avec le solvant attendue pour une pelote statistique divisées par la valeur expérimentale mesurée par RMN.

La relaxation des azotes amides permet de sonder la dynamique rapide du squelette peptidique. On mesure la relaxation longitudinale R_1, la relaxation transverse R_2 et les nOes hétéronucleaires σ. Ces paramètres de relaxation peuvent être reliés à la fonction de densité spectrale qui reflète les contributions des fluctuations dynamiques des vecteurs HN amides en fonction de leurs temps caractéristiques. L'analyse des données de relaxation s'appuie sur différents modèles : l'approche "*model-free*", l'utilisation d'une distribution lorentzienne de temps de corrélation pour décrire la diversité des fluctuations rapides et un modèle dérivé des polymères où les mouvements dominants sont des rotations autour des liaisons peptidiques se propageant le long de la chaine.

Les expériences de dispersion de relaxation permettent de caractériser les phénomènes d'échange conformationnel ou chimique entre deux formes de déplacements chimiques différents pour une gamme de temps allant de la micro à la milliseconde. L'échange dans cette gamme de temps contribue à l'augmentation de la relaxation transverse et provoque un élargissement de raie et une baisse d'intensité pour les pics des résidus impliqués dans l'échange. Dans l'expérience de dispersion de relaxation, des pulses 180°de refocalisation sont appliqués à des fréquences croissantes pendant un délai constant. L'augmentation de la fréquence des pulses de refocalisation entraine une réduction de la contribution de l'échange dans la relaxation transverse. Pour chaque résidu, on analyse la courbe de relaxation transverse apparente en fonction de la fréquence des pulses de refocalisation qui ne montre pas de variation en l'absence d'échange et est globalement décroissante en présence d'échange dans la bonne gamme de temps. Ces profils dépendent de la vitesse d'échange entre les deux états, de la population de l'état minoritaire et de la différence de déplacement chimique entre les deux états.

Les expériences de transfert de saturation (CEST ou Chemical Exchange Saturation Transfer) permettent de caractériser l'échange conformationnel ou chimique entre deux formes de déplacements chimiques différents pour une gamme de temps allant de la milliseconde à la cinquantaine de milliseconde. Le principe de l'expérience est le suivant : pendant une période d'échange, l'échantillon est irradié avec un champ B_1 faible à des offsets réguliers en azote. Si cet offset correspond au déplacement chimique d'un résidu, celui ci disparait du spectre HSQC. Mais si cet offset correspond au déplacement chimique d'une espèce minoritaire en échange, on observe également une baisse d'intensité sur le pic de l'espèce majoritaire. Les profils d'intensités des pics en fonction de l'offset dépendent des déplacements chimiques des espèces majoritaires et minoritaires, de la vitesse d'échange entre les deux états, de la population de l'état minoritaire et des paramètres de relaxation longitudinale et transverse des deux états. Les profils pour chaque résidu sont analysés et ajustés à un modèle théorique d'échange à deux états.

Résultats

Actine recombinante

Ce chapitre présente les résultats obtenus sur l'actine recombinante produite en système baculovirus/Sf9 : production, caractérisation des formes sauvages et mutées et marquage isotopique en cellules d'insectes.

Les actines recombinantes sauvages et mutées ont été produites et purifiées avec succès. Le rendement après gel filtration est de l'ordre de 5mg par litre de culture. La stabilité de ces deux actines recombinantes a été évaluée par thermal shift assay. Les deux actines ont une température de dénaturation identique de l'ordre de 62°C. Cette valeur est du même ordre que celle pour pour l'actine α du muscle squelettique de lapin déterminée par d'autres méthodes.

Pour valider l'utilisation de ces actines recombinantes pour des études structurales d'interaction avec des domaines WH2 nous avons comparé l'interaction de la chimère CH2 précédemment étudiée au laboratoire avec l'actine α du muscle squelettique de lapin et avec les actines recombinantes. L'affinité de CH2 pour ces différentes actines a été évaluée par des expériences de MST. L'ajout d'une quantité croissante de CH2 à une solution d'actine de concentration constante entraine une augmentation de la fluorescence. Les courbes de titration obtenues ont permis la détermination des constantes d'affinité de CH2 pour l'actine de lapin ($K_D = 104 \pm 29$nM) pour l'actine 5C ($K_D = 124 \pm 32$nM) et pour l'actine AP ($K_D = 499 \pm 119$nM). Les courbes de thermophorèse ont une allure plus complexe et n'ont pas permis d'extraire une constante d'affinité. Ainsi, CH2 interagit de manière similaire avec l'actine de lapin et l'actine 5C recombinante mais l'affinité pour l'actine AP est 4 à 5 fois plus faible.

L'interaction de CH2 avec les différentes actines a également été analysée par RMN en comparant les spectres HSQC de corrélation proton azote de CH2 marquée à l'azote 15, en présence d'actine de lapin et en présence des actines recombinantes 5C et AP. CH2 interagit de manière similaire avec l'actine de lapin et avec l'actine 5C. Des petites différences de déplacements chimiques peuvent etre attribuées aux variations de séquence entre les deux actines. En revanche le mode de liaison à l'actine AP est différent, l'hélice amphipatique N-terminale se replie au contact de l'actine AP mais la partie C-terminale reste désordonnée et les pics de cette région sont superposables avec la forme libre de CH2.

Les mutations introduites dans l'actine AP (A204E et P243K) sont situées dans le sous domaine 4 de l'actine, à proximité de la zone d'interaction avec l'hélice C-terminale de CH2. La modification de la taille et de la nature de ces résidus déstabilise l'interaction avec l'hélice C-terminale qui reste désordonnée en présence de l'actine AP. Cette déstabilisation de l'hélice C-terminale est aussi liée à une baisse modérée de l'affinité qui reste principalement gouvernée par la longueur de l'hélice N-terminale et par les interactions électrostatiques intermoléculaire fortes dans la région centrale.

La dernière partie de ce chapitre décrit le développement d'une méthode simple, robuste et peu couteuse pour marquer les protéines exprimées en cellules d'insecte à l'azote 15. La méthode est basée sur l'utilisation d'un milieu commercial sans acides aminés et où la source d'acides aminés est un autolysat de levure enrichi en azote 15 développé à cet effet en partenariat avec un industriel, Biospringer. Cet autolysat contient les 20 acides aminés en quantité suffisante et un traitement thermique a été ajouté pour inactiver les protéases de l'extrait afin d'éviter la protéolyse de protéines sécrétées. Du chlorure d'ammonium est également ajouté ce qui permet d'induire un switch métabolique limitant la consommation de glutamine dans le métabolisme énergétique. Les cellules sont transférées dans le milieu de marquage sans période d'adaptation après infection par le baculovirus recombinant. Il a ainsi été possible de produire de l'actine AP non polymérisable et marquée à l'azote 15 avec un taux d'incorporation de 63% et d'obtenir un spectre HSQC de corrélation proton azote de l'actine AP. Ces premiers résultats sont encourageants mais en raison du haut poids moléculaire de l'actine pour la RMN, il est nécessaire de deutérer au moins partiellement l'actine avant de pouvoir envisager une étude plus poussée.

Caractérisation structurale et dynamique de domaines WH2 libres

Ce chapitre présente la caractérisation structurale et dynamique de 4 domaines WH2 intrinsèquement désordonnés en l'absence d'actine.

Les squelettes peptidiques de CoblAB et NWASPV2 ont été attribués à 278K et pH=7 pour limiter l'échange des protons amides avec le solvant.

On cherche d'abord à savoir si les structures secondaires observées pour les formes liées à l'actine des quatres domaines WH2 étudiés (CH1, CH2, CoblAB et NWASPV2) sont déjà formées de manière transitoire en l'absence d'actine. Pour CH1 et CH2, la forme liée à l'actine contient une hélice α amphipathique N-terminale, un tour d'hélice 3-10 dans le linker et une hélice α C-terminale. Dans les structures cristallographiques, le bidomaine NWASPV2 forme deux hélices amphipathiques une dans chaque domaine. Le bidomaine CoblAB forme deux hélices amphipathiques et une troisième hélice dans la région C-terminale du domaine B.

Les structures secondaires résiduelles ont été évaluées grâce aux déplacements chimiques en utilisant SSP et δ2d. Globalement, les populations des hélices résiduelles évaluées par SSP sont plus importantes que celles évaluées par δ2d. Les hélices amphipathiques sont toutes partiellement préformées mais avec des proportions très variables entre les protéines et entre les différents domaines d'une même protéine, de 8% pour l'hélice amphipathique de CH2 à 42% pour l'hélice amphipathique du domaine B de CoblAB. Pour CH1 et CH2, le tour d'hélice 3-10 dans le linker est également formé de manière transitoire chez les protéines libres en revanche la région C-terminale est complètement désordonnée. Ces deux peptides de séquence proche et de fonctions opposées ont des structures

secondaires résiduelles très similaires. L'hélice C-terminale du domaine B de CoblAB est partiellement formée dans le domaine libre mais elle est plus longue que l'hélice observée dans la structure cristallographique du complexe avec les deux mutants d'actine.

L'analyse des couplages dipolaires résiduels confirme les propensions à former des structures secondaires évaluées par l'analyse des déplacements chimiques. Les valeurs plus élevées qu'attendu pour les couplages dipolaires HN entre les deux hélices amphipathiques de NWASPV2 pourraient être le signe de contacts à longue distance entre ces deux hélices, possiblement favorisés par des interactions hydrophobes. Pour CoblAB, les RDCs HN au sein des hélices ne sont pas totalement conformes aux valeurs attendues, en particulier pour les deux dernières hélices ce qui pourrait indiquer des mouvements à des échelles de temps différentes le long de ces hélices.

La dynamique dans la gamme de la picoseconde à la nanoseconde est analysée à l'aide des expériences de relaxation de l'azote. Pour CH1 et CH2 les expériences de R_1, R_2 et nOes ont été enregistrées à trois champs, 600, 800 et 950MHz afin de mieux décrire la fonction de densité spectrale des mouvements et fluctuations des liaisons HN. Les courbes de R_2 ont globalement une forme en cloche avec quelques résidus présentant des valeurs plus élevées, en particulier les glutamines 43 et 46 dans la région de l'hélice C-terminale. Les paramètres obtenus ont été analysés à l'aide de différents modèles et deux d'entre eux permettent de décrire correctement les données expérimentales. Le premier est un modèle dérivé de l'approche *model-free*. A chaque résidu est associé un temps de corrélation pour les mouvements dans la gamme de la nanoseconde τ_C, un paramètre d'ordre S^2 pour quantifier les mouvements plus rapide dans la gamme de la picoseconde et deux paramètres d'échange $R_{ex} = A_{ex} \times B_0^{\alpha}$. Les valeurs de α sont autour de 1 ce qui correspond à de l'échange intermédiaire à l'échelle des déplacements chimiques. Le deuxième est un modèle dérivé de l'analyse des polymères où le mouvement dominant en solution est un mouvement de rotation autour des liaisons covalentes du squelette qui se propage de la chaine peptidique. Ces mouvements sont caractérisés par un temps caractéristique de la transition τ_1, un temps caractéristique pour l'atténuation de la propagation du mouvement le long de la chaine peptidique τ_2 et un paramètre d'ordre A_{HH} analogue au S^2 de l'approche *model-free*.

Les paramètres de relaxation de NWASPV2 et de CoblAB n'ont été mesurés qu'à un champ. L'analyse est donc plus qualitative. Pour NWASPV2, les profils de R_1 et nOes sont globalement plats avec peu de variations. Les deux régions contenant les hélices transitoires ont des valeurs de R_2 plus élevées. Deux arginines, en région C-terminale de chacune des hélices ont des valeurs de R_2 plus élevées. Ces résidus sont conservées dans de nombreux domaines WH2 comme ceux des protéines WASP, WASH, WAVE, MIM et certains des domaines WH2 de spire et Cordon-bleu et interagissent avec la thréonine 148 de l'actine (PDB : 2VCP). Cordon-bleu pour lequel les hélices transitoires sont formées avec des proportions plus importantes présente des profils de R_1 et nOes avec des variations plus importantes et une nette diminution des R_1 et augmentation des nOes dans les ré-

gions formant les hélices transitoires ce qui est cohérent avec une rigidité plus importante dans ces régions. Une très nette augmentation de la relaxation est également observée dans ces régions en particulier pour les deuxièmes moitiés des hélices amphipathiques.

Les mouvements dans la gamme de la microseconde à la milliseconde ont été analysés grâce à des expériences de dispersion de relaxation et d'échange des protons amides avec le solvant.

Pour CH1 et CH2, les expériences de dispersion de relaxation montrent qu'il existe de l'échange tout le long de la chaine peptidique, avec des variations le long de la chaine, et sans qu'il y ait de zone d'échange plus important clairement identifiable. L'échange mesuré par dispersion de relaxation est lié à un changement d'environnement du noyau considéré qui peut être du soit à des changements conformationnels, soit à de l'échange chimique des protons amides avec les protons du solvant. Une expérience de contrôle de mesure des vitesses d'échange des protons amides avec le solvant a permis de montrer que la dispersion mesurée correspond plutôt à de l'échange avec le solvant. Les facteurs de protections sont globalement inférieurs à 1, sauf, étonnamment, dans la deuxième moité de l'hélice C-terminale qui est complètement désordonnée en l'absence d'actine d'après l'analyse des déplacements chimiques et des couplages dipolaires résiduels. Il est possible que des interactions électrostatiques non spécifiques dans cette région très chargée rendent le squelette peptidique moins accessible.

Pour NWASPV2, les expériences de dispersion de relaxation montrent qu'il existe de l'échange tout le long de la chaine peptidique. Les valeurs mesurées sont trop élevées pour correspondre à de l'échange chimique avec le solvant et correspondent vraisemblablement plutôt à de l'échange conformationnel. Des paramètres d'échanges similaires ont été obtenus pour les deux domaines ce qui semble indiquer que la transition implique la protéine dans son ensemble. Il est possible que l'échange mesuré corresponde à un échange entre une conformation, ou un ensemble de conformations où les deux hélices amphipatiques interagissent entre elles et une conformation plus étendue, des contacts à longue distance ayant déjà été suggérés par l'analyse des expériences de couplages dipolaires résiduels. Les facteurs de protection sont globalement très élevés, en accord avec une chaine peptidique rendue peu accessible par des contacts à longue distance.

Pour CoblAB, la comparaison de la dispersion de relaxation expérimentale avec la contribution liée à l'échange avec le solvant montre qu'il existe de la dispersion liée à l'échange conformationnel localisée sur quelques résidus dans la deuxième moitié des hélices amphipathiques. L'ajustement des courbes à un modèle à deux états pour ces deux régions séparément conduit à des populations pour l'état minoritaire de 1 à 2% et des temps caractéristiques d'échange de 1 à 2 millisecondes pour la première hélice et 5 à 10 millisecondes pour la deuxième. Cet échange pourrait correspondre à une transition entre les formes dépliées et repliées dans les régions des hélices transitoires. Les facteurs de protection pour CoblAB sont globalement inférieurs à 1, sauf dans les deuxièmes moitiés d'hélices et dans l'étiquette Strep. Il est intéressant de noter que les facteurs de protection

sont plus élevés pour la deuxième hélice amphipathique en accord avec un temps de vie plus long des conformères hélicoïdaux dans cette région.

Interaction avec l'actine

L'interaction entre CH1 et l'actine 5C est suivie par des expériences de CEST pour différentes concentrations en sel. L'expérience de CEST-N15 développée récemment (Vallurupalli et al. 2012) a été implémentée sur nos spectromètres et a été validée sur une protéine modèle du laboratoire.

A première vue, les profils correspondent à un échange à deux états et les caractéristiques de l'échange dépendent de la concentration en sel comme précédemment observé. Les déplacements chimiques de l'état minoritaire correspondent à ceux de la forme complexée à l'actine précédemment caractérisée.

L'ajustement quantitatif des courbes se fait par cluster pour trois régions distinctes (l'hélice N-terminale, la région centrale et l'hélice C-terminale) et pour l'ensemble de la protéine. Globalement, la vitesse d'échange augmente avec la concentration en sel.

A 80mM en sel, la vitesse d'échange déterminée pour l'hélice C-terminale est plus grande que pour les autres régions, en accord avec un décrochage de cette hélice et une augmentation de la dynamique dans la région C-terminale. Il est intéressant de noter que les paramètres diffèrent selon le choix du cluster à l'intérieur de cette hélice. Les valeurs de p_B sont plus grandes et les valeurs de k_{ex} plus faibles pour les clusters de résidus en N-terminal de cette hélice et les valeurs de p_B augmentent et les valeurs de k_{ex} diminuent pour les clusters de résidus en C-terminal de l'hélice. Cela semble indique une dynamique non uniforme le long de l'hélice et un décrochage plus marqué en région C-terminale. Ce phénomène n'est plus observé à 100mM en KCl. Pour cette condition, les valeurs de p_B et k_{ex} pour différents clusters de l'hélice C-terminale sont homogènes en accord avec un décrochage plus uniforme de l'hélice.

L'interaction entre NWASPV2 et l'actine est suivie par RMN en analysant les spectres (^{1}H-^{15}N)-HSQC de NWASP en présence de quantités croissantes d'actine. D'après les expériences de titrage en fluorescence, on s'attend à une interaction non coopérative avec une affinité significativement différente pour chacun des deux domaines WH2. Dans ce cas, un complexe de stœchiométrie 1:1 devrait être majoritaire jusqu'à des ratios molaires actine:NWASPV2 de 1:1, le deuxième site de fixation n'étant occupé que pour des ratios d'actine plus élevés.

L'ajout d'actine à la solution de NWASPV2 se traduit par une baisse d'intensité des résidus des deux domaines WH2. On ne se trouve donc pas dans un cas d'interaction non coopérative avec deux sites d'affinités différentes mais plutôt dans le cas d'une interaction coopérative où un complexe de stœchiométrie 1:2 est prédominant même en défaut d'actine et est de taille trop importante pour être observé sur nos spectres. Une autre possibilité est une interaction non coopérative avec deux sites d'affinités équivalentes.

Ces expériences ont permis de définir plus précisément les zones d'interaction avec l'actine. Les douze derniers résidus de notre construction interagissent plus faiblement avec l'actine car la baisse d'intensité est plus faible dans cette région. La déstabilisation de l'interaction de la région C-terminale du deuxième domaine WH2 avec l'actine peut s'expliquer par le remplacement d'une arginine présente dans le premier domaine et formant un pont salin avec un glutamate de l'actine, par un glutamate dans le deuxième domaine.

L'interaction entre CoblAB et l'actine est suivie par RMN en analysant les spectres (^{1}H-^{15}N)-HSQC de CoblAB en présence de quantités croissantes d'actine. CoblAB est triplement marqué pour optimiser les propriétés de relaxation afin d'espérer observer directement le complexe.

L'ajout d'actine à la solution de NWASPV2 se traduit par une baisse d'intensité des pics et par l'apparition de quelques nouveaux pics, on se trouve donc dans un cas d'échange lent. L'analyse des profils de baisse d'intensité lors de l'ajout d'actine montre que l'actine interagit préférentiellement avec le domaine B de CoblAB et que la région entre l'hélice amphipathique du domaine A et l'hélice amphipathique du domaine B forme un long linker flexible qui interagit plus faiblement avec l'actine.

Une partie des pics qui apparaissent lors de l'ajout d'actine ont été attribués et correspondent à la région C-terminale du domaine B, des résidus 88 à 100. L'analyse des différences de déplacement chimique entre les formes libres et liées montre la formation d'une hélice α pour les résidus 87 à 90. Il n'est malheureusement pas possible d'observer complètement la formation d'une hélice longue dans le complexe en solution allant des résidus 73 à 90 et correspondant à l'hélice transitoire de la forme libre.

Conclusion générale

Cette thèse s'inscrit dans un projet visant à comprendre le remodelage du cytosquelette d'actine par les domaines WH2 et comment des interactions spécifiques entre protéines hautement dynamiques permettent d'exercer de nombreuses fonctions.

Nous avons mis en place la production d'actine recombinante pour les études structurales. Grâce au développement de nouvelles méthodes de marquage, nous avons pu produire de l'actine marquée à l'azote 15 pour la RMN ce qui constitue un premier pas pour la caractérisation en solution des états multiples de l'actine par RMN.

Nous avons étudié les ensembles conformationnels de plusieurs domaines WH2 isolés et bidomaines. Les protéines étudiées forment des ensembles hautement dynamiques comportant des hélices résiduelles dans des proportions variables. La dynamique de ces domaines semble également plutôt variable avec une dominante de mouvements très rapides dans la gamme des nanosecondes pour certains domaines et une dynamique plus lente, de la micro à la milliseconde sur l'ensemble de la protéine ou sur des régions localisées pour d'autres domaines.

Enfin, nous avons pu caractériser l'interaction de deux bidomaines WH2 avec l'actine. Bien qu'il soit difficile de caractériser le complexe directement, les expériences de titrage en RMN ont permis de délimiter précisément les zones d'interaction et de mieux comprendre l'interaction de ces bidomaines avec deux actines en terme de coopérativité, différence d'affinité des deux sites, flexibilité locale, etc. Pour observer le complexe lié à deux actines, la deutération du peptide ne suffit pas comme nous l'avons vu dans le cas de CoblAB. Il faudra envisager la deutération au moins partielle de l'actine ou des marquages spécifiques de méthyles du bidomaine WH2 pour espérer observer le complexe directement par RMN. Une autre solution est de caractériser le complexe à partir de l'échange avec la forme libre. Cela devrait être possible grâce à des expériences de CEST déjà utilisées pour analyser l'interaction de l'actine avec un seul domaine WH2.

I

Biological context

Intrinsically disordered proteins 1

1.1 Re-assessing the protein structure-function paradigm

For a large part of the 20th century, structural biology was governed by the structure-function paradigm based on the idea that a defined three-dimensional structure is required for a protein to be functional. Based on analysis of catalytic processes in enzymes, Fisher first proposed a "lock and key" model where a binding site only accommodates substrates of particular three-dimensional shape explaining specificity of protein functions (Fisher 1894). Koshland (1958) then suggested that protein dynamics should be taken into account to explain the structural variety of substrates binding some enzymes : this is the "induced-fit" model.

In addition to these folded, more or less dynamic proteins, peptides, hormones but also large protein regions and even full proteins lack stable three-dimensional structure in native conditions but are functional (Wright and Dyson 1999). Intrinsically disordered proteins or regions (IDPs or IDRs) are far from anecdotal. About 30% of eucaryotic proteins and 44% of the proteins of the human genome are predicted to contain disordered segments of more than 30 amino acids (Ward et al. 2004, Oates et al. 2013, van der Lee et al. 2014).

Disorder is encoded in the amino acid sequence and IDPs or IDRs are characterized by low proportions of hydrophobic, order-promoting, residues and high proportions of charged residues, glycines and prolines, low sequence complexity and repetitiveness (Tompa 2003). Many IDPs are polyampholytes and their conformational preferences are determined by charge distribution along the sequence and in particular distribution of oppositely charged residues. Sequences with alternating positive and negative residues are expected to prefer more extended conformations whereas sequences with clustered positive and negative residues are expected to adopt more compact conformations (Das and Pappu 2013).

A variety of methods were developed to predict disorder from an amino acid sequence based on prediction of the physico-chemical properties of the peptide chain depending on the amino acid sequence like IUPred (Dosztanyi et al. 2005) or machine-learning algorithms which integrate known properties characterized experimentally like PONDR XL1 (Romero et al. 2004). A simple way is to plot the mean net charge as a function of mean

normalized hydrophobicity. Intrinsically disordered proteins are on the low mean hydrophobicity, high net charge region of this CH plot or Uversky plot (Dunker et al. 2008), separated from folded proteins by a line of equation :

$$\langle R \rangle = 2.743\langle H \rangle - 1.109 \tag{1.1}$$

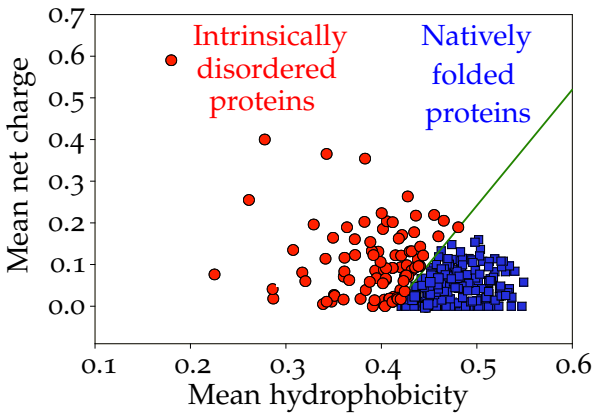


Figure 1.1 – *Charge-hydropathy plot or Uversky plot showing mean net charge as a function of mean normalized hydrophobicity for natively folded or disordered proteins (Dunker et al. 2008)*

This is however a simplified view of and there is actually no clear frontier between order and disorder but rather a continuum of possibilities from a folded and rigid protein to an extended random coil one as illustrated on figure 1.2.

Figure 1.2 – *Continuum of disorder from maximum order on the left to complete disorder on the right : folded protein with no disordered region, disordered N- and/or C-terminal regions, disordered loop or linkers, disordered domains, disordered protein with residual structure, completely disordered proteins with compact or extended conformations (Uversky et al. 2005, Habchi et al. 2014).*

IDPs lack stable three dimensional structure so they are better described by conformational ensembles that reflect their structural diversity (Varadi et al. 2014). Structural characterization of the conformational preferences, residual structure and long range contacts in IDPs can help to understand their functions and modes of action. Classically used experimental methods include circular dichroism, vibrational spectroscopy, SAXS, EPR or fluorescence (Uversky and Dunker 2012). Nuclear magnetic resonance spectroscopy is a method of choice to study structure and dynamics of IDPs at atomic resolution. NMR-based methods are further described in the experimental section.

These proteins are not intrinsically disordered only because of the experimental conditions of their characterization *in vitro* and intrinsic disorder is relevant *in vivo*. Effects of macromolecular crowing depend on the protein and the crowing agent. Some IDPs respond to molecular crowding by different degrees of compaction (Soranno et al. 2014)

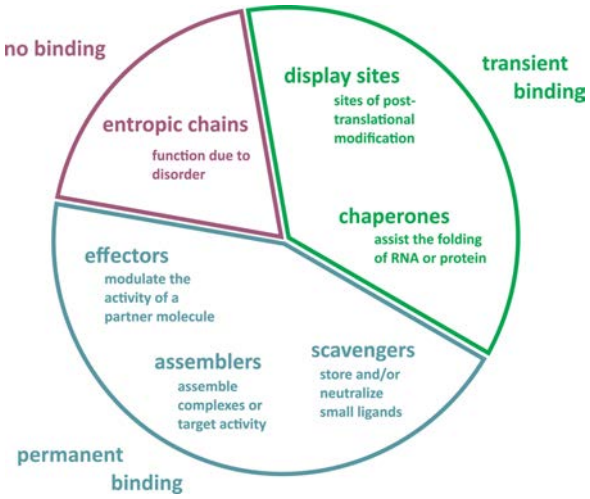


Figure 1.3 – *Functions of intrinsically disordered proteins or regions (van der Lee et al. 2014).*

whereas other are not affected (Cino et al. 2012). In-cell NMR experiments on several IDPs like α-synuclein (McNulty et al. 2006) or thymosin-β_4 (Ogino et al. 2009) show that they are still disordered in the cellular environment.

1.2 Functions of intrinsically disordered proteins

As they are very abundant in genomes of eucaryotes, intrinsically disordered proteins or regions are involved in various cellular processes including signaling, regulation of transcription, cytoskeletal dynamics and remodeling etc. Being involved in many regulatory processes, they are linked to various diseases such as cancers.

Functions of intrinsically disordered proteins or regions can be classified in 6 general categories (Tompa 2005, van der Lee et al. 2014) summarized in figure 1.3.

The first class of IDP/IDRs is composed of entropic chains including flexible linkers and entropic springs. In that case, the function is directly related to the physical properties of the peptide chain and does not require partner binding or order gain. For example, titin is the longest human protein. Very abundant in muscle cells it is responsible for the passive elasticity of muscles via its PEVK regions enriched in these four amino acids (Gutierrez-Cruz et al. 2001). It was proposed that the elasticity of the PEVK region and preferred conformations were finely tuned by the charge repartition in this polyampholyte chain (Forbes et al. 2005).

Other IDP/IDRs functions are linked with transient or permanent binding with partners.

IDRs may serve as display sites for post-translational modifications. Intrinsically disordered regions are generally enriched in sites for post translational modifications including phosphorylation (Gsponer et al. 2008), methylation, acetylation... Intrinsic disorder enhances accessibility of the site to modify for binding of the modifying enzyme and of effectors proteins for signal transmission (van der Lee et al. 2014, Galea et al. 2008). Post

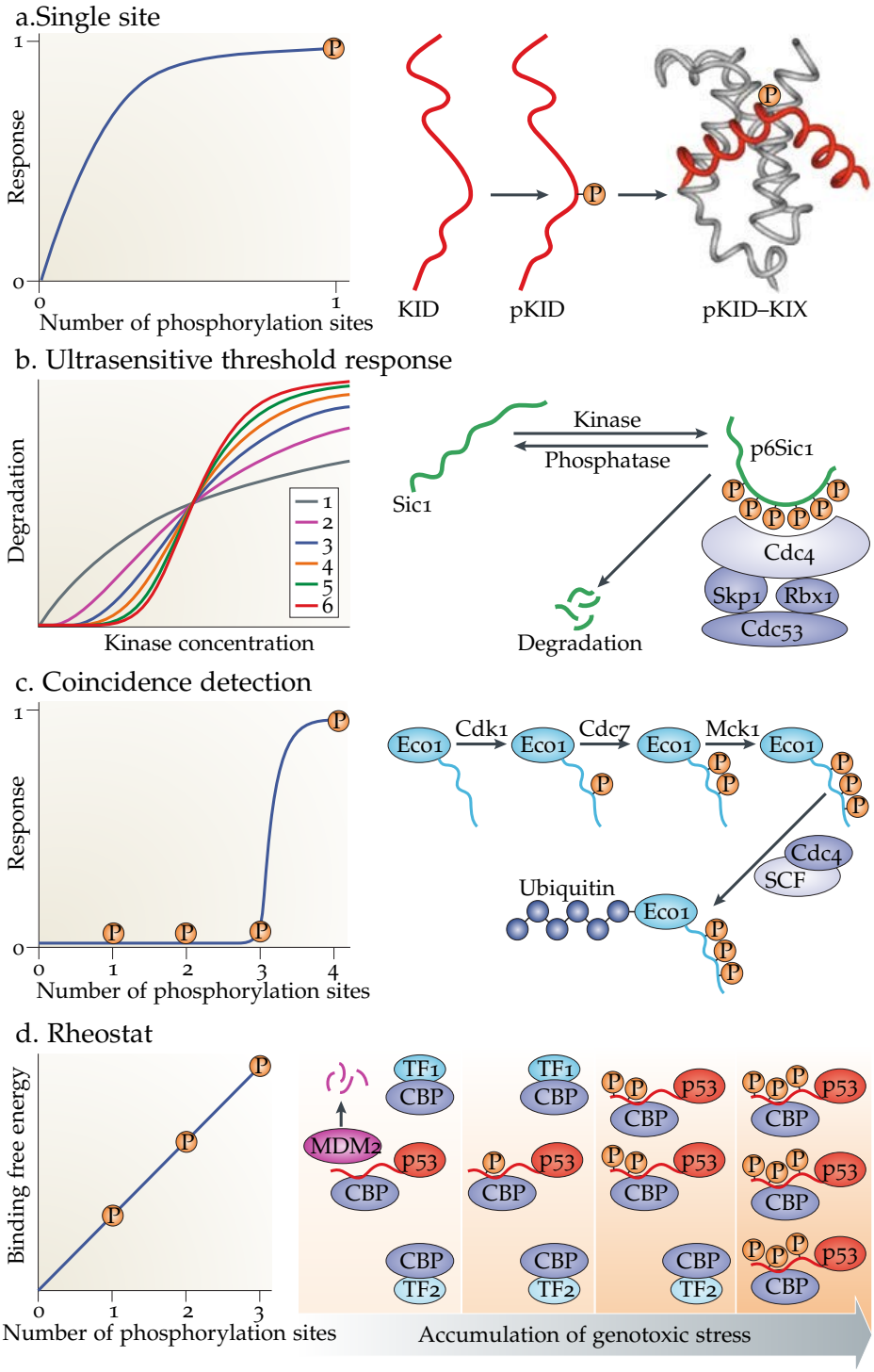


Figure 1.4 – *Possible responses to multisite phosphorylation in IDPs involved in signalling (Wright and Dyson 2015).*

translational modifications in disordered histone tails are crucial in gene transcription regulation (Hansen et al. 2006). Combinatorial post-translational modifications of intrinsically disordered regulatory proteins allow interaction of complex signaling processes (Wright and Dyson 2015). A few examples of possible responses to multisite phosphorylation in IDPs are illustrated on figure 1.4.

Intrinsic disorder also plays an important role in chaperones which assist protein or RNA folding as disordered regions represent half of the sequences of RNA chaperones and a third of the sequences of protein chaperones.

Some intrinsically disordered proteins act as effectors and interact with other proteins to modulate their activity. Interaction may be accompanied by partial or global folding of the intrinsically disordered protein (Wright and Dyson 2009). Characteristics of protein-protein interactions involving IDPs will be detailed in the following section 1.3. IDPs are also involved in auto-regulatory processes within the same protein. The role of intrinsic disorder in auto-inhibition and activation of the WASP/N-WASP proteins will be detailed in section **??**.

Disordered assemblers recruits different partners and enable scaffolding of multiprotein complexes (Fuxreiter et al. 2014). They provide more binding interface than folded proteins for the same number of amino acids. Several binding sites such as molecular recognition features (MoRFs), short linear peptide motifs (SLiMs) are distributed along the sequence so these disordered assemblers form interaction hubs.

Another class of IDP/IDRs are scavengers which bind and, and neutralize small ligands. Secreted calcium phosphate-binding proteins (SCPPs) are mostly disordered and sequester calcium-phosphate clusters in calcium hypersaturated biofluids like milk (Holt 2013).

Finally, some IDPs of low sequence complexity are involved in phase transitions to form droplets or granules which constitute membrane-less organelles in the cytoplasm (Kato et al. 2012, Weber and Brangwynne 2012).

1.3 Protein-protein interactions involving IDPs

Many IDPs operate through transient or permanent protein-protein interactions. IDPs interactions with their partner are generally characterized by moderate affinities (in the μM range) and high specificity, large interaction surfaces, conformational plasticity and adaptability and are often mediated by rather short molecular recognition elements (MoREs) or molecular recognition features (MoRFs) (Oldfield et al. 2005).

1.3.1 Folding upon binding

Binding of these MoREs/MoRFs is often accompanied by disorder to order transitions leading to α-helices, β-strands or irregular secondary structures in the bound form. Comparison of protein-protein interactions involving only folded proteins on the one hand and

at least one IDP or IDR on the other hand showed that the latter contain decreased proportions of β-strand conformations and increased proportions of coil structures (Habchi et al. 2014). High representation of helical structures in IDPs interactions is understandable because helix formation is local and only implies medium range interactions thus facilitating the folding process compared to structures requiring long range intramolecular contacts. These more ordered complexes may be amenable to structural studies for conventional X-ray crystallography or NMR structure determination.

A particular interesting feature of protein-protein interactions involving IDPs is that specificity is decoupled from binding affinity (Oldfield et al. 2005). Indeed, interaction surfaces are high because of extended nature of IDPs leading to important specificity but the entropic cost for binding is important thus decreasing the binding affinity. It allows highly specific reversible interactions particularly relevant in regulatory processes.

Many IDPs contain preformed structural elements and the free protein may have conformational preferences for secondary structures present in the bound state (Fuxreiter et al. 2004). Therefore analysis of free IDPs can provide first insights into interacting regions and possible nature of the interaction.

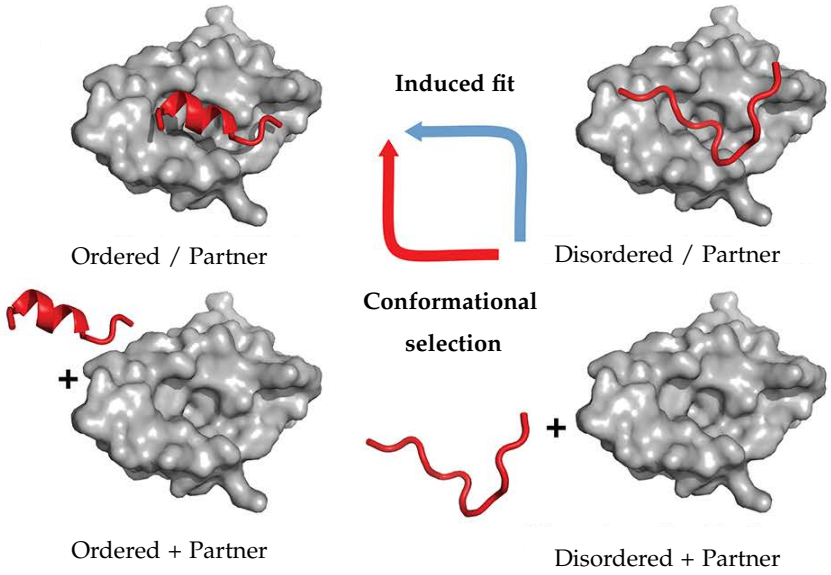


Figure 1.5 – *Representation of coupled folding upon binding via conformational selection or induced fit mechanisms (Theillet et al. 2014).*

The disorder to order transition in folding upon binding processes can be described by two extreme mechanisms (Wright and Dyson 2009). If the IDP contains preformed structural elements, one possibility is that only folded conformers interact with the partner. In that case, folding happens before binding. This is the conformational selection mechanism whose benefits is lowering of the entropic binding cost. Another possibility is that the IDP interacts before folding and folding is driven by intermolecular interactions with the partner's interface. This is the induced fit mechanism which puts forward IDP conformational adaptability to its binding partner. A third proposed interaction mechanism is

fly-casting or "protein fishing" in which IDPs take advantage of their high capture radius to probe their environment and search for partners (Shoemaker et al. 2000). It is likely that many folding upon binding processes may not be described by a single mechanism but by a combination of features of conformational selection, induced fit and fly-casting (Espinoza-Fonseca 2009).

1.3.2 Dynamics and adaptability of bound IDPs

Very often, IDPs that fold upon binding retain dynamics in the bound state thus forming fuzzy complexes with their interacting partner (Tompa and Fuxreiter 2008). Four manifestation of fuzzyness are represented on figure 1.6. WH2 repeats presented in chapter 3 form fuzzy complexes close to the flanking model when they interact with actin.

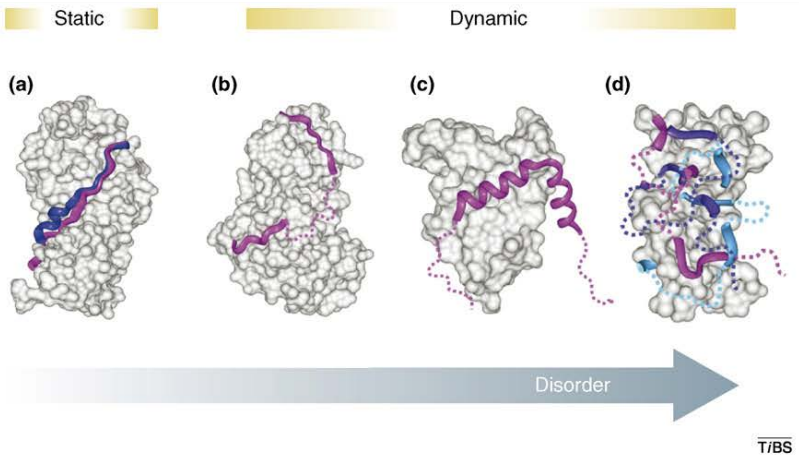


Figure 1.6 – *Examples of fuzzy complexes (Tompa and Fuxreiter 2008). a. Polymorphic interaction : several conformations are populated, involving different intermolecular interactions, illustrated by the interaction of Tcf4 with β-catenin (PDB : 1JDH). b. Clamp model : two folded molecular recognition features are connected by a flexible disorder linker, illustrated by the interaction of Ste5p with the MAP kinase Fus3p. c. Flanking model : regions adjacent to the molecular recognition feature interact with the partner in a dynamics manner and are not resolved on the structure, illustrated by interaction of CREB KID with CBP KIX (PDB : 1KDX). d. Random interaction : the IDP remains disordered in the bound state and binds its partner via transient interactions, illustrated by a model of T cell receptor ζ-chains.*

So far, we described interactions of a single IDP with one partner but thanks to their conformational plasticity and adaptability, some IDPs are able to interact with different partners, using a single molecular recognition element. That is the case of the C-terminal region of the tumor suppressor p53 that interacts with different proteins and adopts various helical, extended and coil conformations in the bound state as illustrated on figure 1.7 (Oldfield et al. 2008). Besides, different intrinsically disordered proteins may interact with the same partner to modify its activity. This is the case of WH2 repeats which regulate actin polymerization into filaments.

Many proteins involved in the regulation of the cytoskeleton dynamics and remodeling are intrinsically disordered or contain intrinsically disordered regions (Guharoy et al.

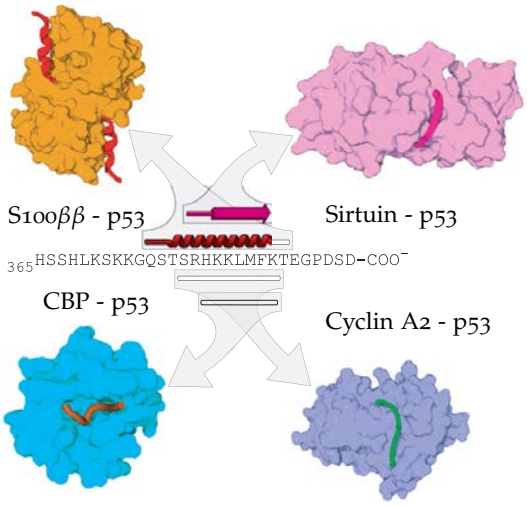


Figure 1.7 – *Adaptability of the C-terminal region of tumor suppressor p53 to binding of various partners (Oldfield et al. 2008).*

2013). The present work will focus on WH2 repeats, intrinsically disordered domains involved in the regulation of the actin cytoskeleton.

Actin, an essential cytoskeletal protein

2

The cytoskeleton is the ensemble of biological polymers which confer its mechanical properties to a cell. It is involved in establishment, maintenance and adaptation of the cell architecture and organisation, cell motility and adhesion, intracellular trafficking and muscular contraction. The cytoskeleton of eucaryotes is composed of three polymers : microtubules, intermediate filaments and actin microfilaments. This chapter will briefly introduce the actin cytoskeleton and its regulation.

2.1 General introduction

Actin, discovered in 1942 by Brunó Ferenc Straub in rabbit muscle extract (Straub 1942) is one of the most abundant cytosolic protein in eucaryotic cells with concentrations above 100μM in many cell types and up to 550μM in human platelets (Pollard et al. 2000). In muscles it represents 20% of the total mass of proteins in the cell (Sheterline et al. 1995).

The actin cytoskeleton is a dynamic network of filaments that takes part in many cellular processes. Actin is involved in muscular contraction via the actomyosin myofibrils (Bagshaw 1992). In the cytoplasm of non muscle cells, the actin cytoskeleton also supports processes such as establishment and maintenance of cell polarity, adaptation of cell shape, cellular motility and adhesion, cytokinesis, endocytosis and intracellular trafficking (Pollard and Cooper 2009). Early orbservations of actin in the nucleus were first controversial but it is now accepted that nuclear actin plays a role in transcription, chromatin remodeling and signal transduction (Hendzel 2014).

Because actin is involved in a myriad of cellular processes, abnormalities and misregulations generally result in diseases and pathologies. Mutations in the actin genes which modify the actin filament properties were linked to congenital myopathies (Clarkson et al. 2004). Actin cytoskeleton misregulation and rearrangements were shown to be involved in tumor cells migration and invasion and in development of metastasis (Nurnberg et al. 2011). Some pathogens agents hijack the actin cytoskeleton of the host during their life cycle such as *Listeria* bacterium which uses a comet of actin filaments to move in the cytoplasm of the infected cell (Tilney et al. 1992a;b) or the HIV which uses the host actin cytoskeleton for entry and replication mechanisms (Stolp and Fackler 2011). This non

exhaustive list of pathologies underlines how crucial it is to understand the molecular mechanisms involved in actin cytoskeleton assembly and regulation.

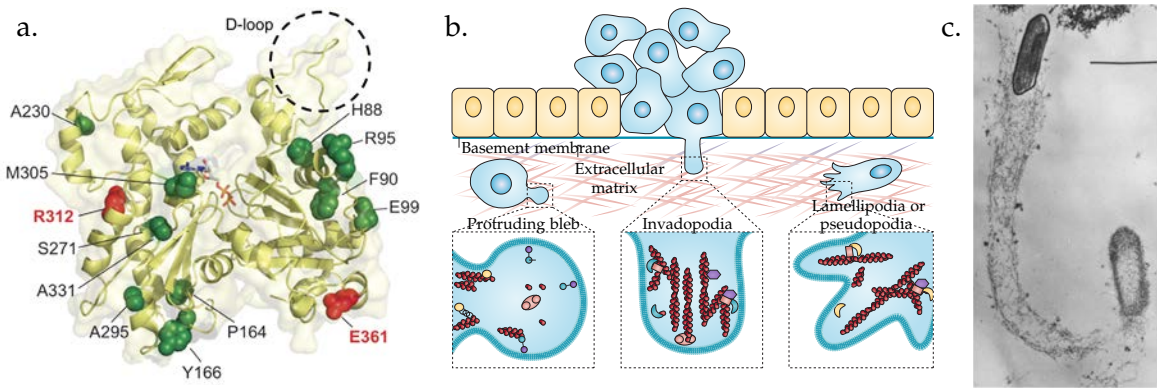


Figure 2.1 – *Examples highlighting actin-linked pathologies. a. Mutations in human α-cardiac actin related to hypertrophic cardiomyopathy in green and to dilated cardiomyopathy in red (Mundia et al. 2012) b. typical protrusive structures in invasive cancer cells (Nurnberg et al. 2011). Protrusive blebs (on the left) generate protrusive force that deforms the plasma membrane of the cell in a particular direction to move through the extracellular matrix. Invadopodia (in the middle) are actin rich protrusions involved in the degradation of the extracellular matrix. Lamellipodia or pseudopodia (on the right) are also actin rich protrusion involved in the cell motility through the extracellular matrix. c. Electron microscopy image of a* Listeria *bacterium with a long tail of actin filaments. The bar corresponds to 1µm (Tilney et al. 1992a)*

2.2 G-actin

2.2.1 Key features

G-actin designates monomeric globular actin, a 43kDa protein comprising 373 to 375 amino acids depending on species and isoforms, present in all eucaryotic cells. Actin sequence is highly conserved with 86% sequence identity between human α-skeletal actin and yeast (*Saccharomyces cerevisiae*) and 100% sequence identity between human, rabbit and chicken α-skeletal actins. Higher mammals like human have six actin isoforms expressed by different genes : α-skeletal, α-cardiac, α-smooth and γ-smooth actins are expressed in muscles and β-cytoplasmic and γ-cytoplasmic actins are ubiquitous. The six human actin isoforms share 92% of sequence identity and a third of the sequence differences are concentrated in the 20 N-terminal residues.

This exceptional sequence conservation was proposed to be caused by the fact that the sequence was optimized to confer the proper mechanical properties to the actin filaments (Galkin et al. 2012) or by the many interactions with over 100-200 actin binding proteins. Using the STRING database which produces networks of predicted associations for a particular group of proteins, 193 interacting proteins are identified with high confidence scores ($\geq$ 0.7) by experimental, co-expression, database or textmining evidences for the sole human α-skeletal actin.

All actins have similar physico-chemical properties among which a low isoelectric

point of 5.5, a high affinity binding site for a nucleotide (ATP or ADP) and a divalent ion (magnesium Mg^{2+} in physiological conditions). Conserved post-translational modifications are cleavage of 1 to 2 N-terminal amino acids, acetylation of the N-terminus and methylation of histidine 71 close to the nucleotide binding site.

2.2.2 Stuctures of the actin monomer

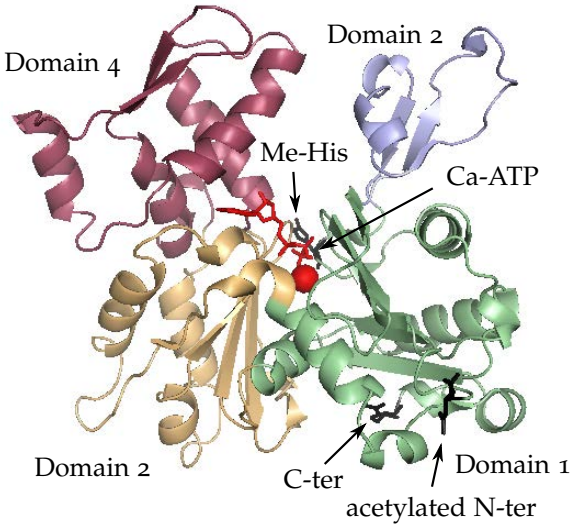
The structure resolution of G-actin is essential to explain actin polymerization into filaments and to characterize regulatory processes by actin binding proteins. The main obstacle for structural investigation is that G-actin spontaneously forms filaments rather than crystals at high concentration. The first G-actin structure was obtained with a 2.8 angtroms resolution from crystals of a non-polymerizable complex with DNase I (Kabsch et al. 1990). Since then, many crystal structures of G-actin were solved, in complex with various actin binding proteins and small molecules (like latrunculin, a marine toxin which inhibits actin polymerization Bubb et al. (2002)). Actin modifications such as cross-linking with tetramethylrhodamine (Otterbein et al. 2001, Graceffa and Dominguez 2003) or point mutations (Rould et al. 2006) enabled crystallization of uncomplexed actin. Crystallization of native unmodified G-actin was achieved in presence of stoechiometric amounts of heat shock protein Hsp27, a chaperone that prevents actin/actin interactions in solution thus inhibiting filament formation and enabling the growth crystals were only G-actin and not Hsp27 was found (Wang et al. 2010). To date, 118 structures of the protein data bank contain a α-skeletal actin molecule.

All of these actin structures are very similar. G-actin has a flattened elliptic shape and consists of two major domains I and II each of which can be divised into 2 subdomains (see figure 2.2). A nucleotide (ATP or ADP) and a divalent ion (Mg^{2+} in physiological conditions or Ca^{2+} in most crystal structure) are bound to a central cleft between domains I and II. An hydrophobic cleft is accessible between subdomains 1 and 3. Because of the polarized nature of actin assembly, actin monomers are sometimes represented as arrows with a barbed end (subdomains 1 and 3) and a pointed end (subdomains 2 and 4).

Significant differences among G-actin structures are observed in localized loop regions : the DNase I binding loop or D-loop (residues 40 to 51 in subdomain 2) which is disordered in some crystal structures and forms a α-helix in others. Other fluctuating regions are the S-loop (residues 11 to 16), the H-loop (residues 70 to 78), the G-loop (residues 154 to 161) and the W-loop (residues 165 to 172) which appear to be more dynamic as attested by their higher B-factors in many crystal structures.

2.3 F-actin

Above a critical concentration which depends on experimental conditions, G-actin polymerizes into actin filaments or F-actin which constitute the actin cytoskeleton. Once



Domain organization	
Major domain I	
Domain 1	residues : 1-32, 70-144, 338-375
Domain 2	residues : 33-69
Major domain II	
Domain 3	residues : 145- 180, 270-337
Domain 4	residues : 181-269

Figure 2.2 – *Domain organization of G-actin : crystal structure of monomeric Ca-ATP actin (PDB : 1ATN). The four domains are highlighted in different colors, the divalent ion and nucleotide are shown as red sphere and sticks and the acetylated N-terminus, the methylated His71 and the C-terminus are shown as black stick.*

formed, these filaments remain highly dynamic and constantly polymerize and depolymerize. *In cellulo*, actin assembly and disassembly is tightly regulated by actin binding proteins to control and coordinate the various above-mentioned cellular processes.

2.3.1 Mechanism of actin polymerization

In vitro actin assembly into filaments proceeds through a activation/nucleation/elongation mechanism when actin concentration exceeds the critical concentration for filament assembly (C_C).

2.3.1.1 Activation

The critical concentration depends on many factors such as temperature, nature of the divalent cation and nucleotide bound to actin and ionic composition of the buffer.

The critical concentration is lowered when actin is bound to a Mg^{2+} cation rather than Ca^{2+} and to ATP rather than ADP (Carlier 1991).

In vitro, actin assembly is induced by a change of buffer, from a low ionic strength G-buffer that stabilizes monomers to a physiological ionic strength F-buffer that induces filament formation. Monovalent (K^+) and divalent (Mg^{2+}) ions interact with actin a low affinity (K_D in the micro to millimolar range) to neutralize charges, lower the energetic cost for intermolecular interactions between actin monomers and induce a conformational change (Rouayrenc and Travers 1981, Pardee and Spudich 1982).

Actin is a negatively charged polyelectrolyte because of its low isoelectric point of 5-5.5. Electrostatic potentials at the surface of rabbit α-skeletal G-actin are represented on figure 2.3. Assembly of this charged biopolymer is expected to be affected by electrostatic

G-buffer	Tris (pH=7.8)		0.1mM $CaCl_2$	0.2mM ATP	1mM DTT
F-buffer	Tris (pH=7.8)	100mM KCl	0.1mM $MgCl_2$	0.2mM ATP	1mM DTT

Table 2.1 – *Composition of the G and F-buffers*

interactions with ions of an electrolyte solution such as the cytoplasm. Non specific interactions screen electrostatic repulsion between the negatively charges monomers. Moreover, specific binding-sites for divalent cations, identified on actin monomers by bioinformatic tools and verified by mutagenesis (Kang et al. 2012), drive actin polymerization.

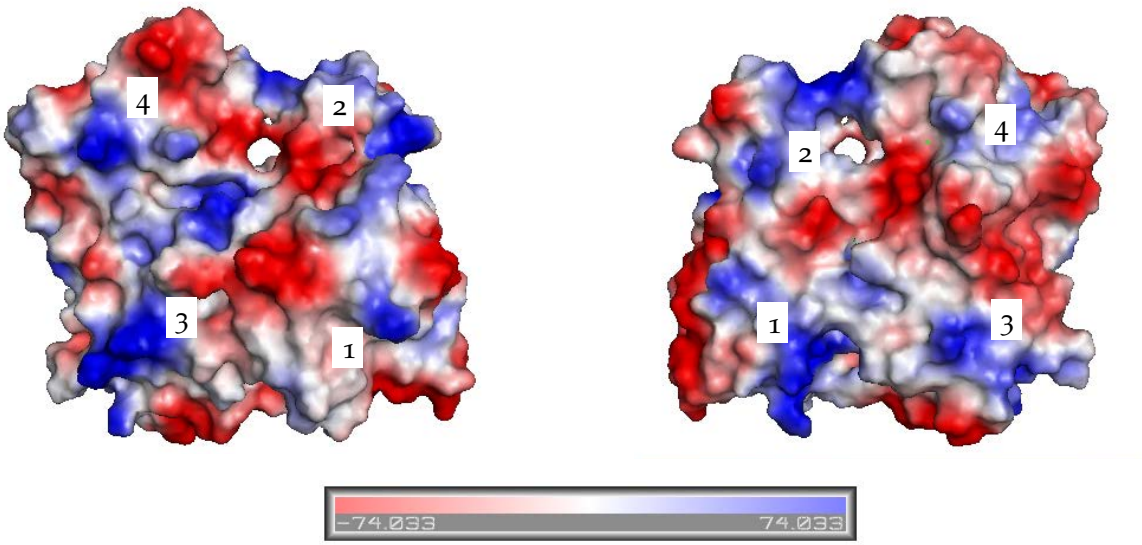


Figure 2.3 – *Electrostatic potentials at the surface of rabbit α-skeletal G-actin (generated using PDB : 3HBT). Positive and negative charges on the surface are represented in blue and red respectively. The side of the protein shown on the left is exposed to the solvent and the side of the protein shown on the right is buried in the filament.*

2.3.1.2 Nucleation

The nucleation step is the formation of actin nuclei, dimers then trimers, which form the basis of the helicoidal structure of the filament. This is the rate-limiting step as determined by computer simulations and free energy calculations taking into account electrostatic and hydrophobic interactions as well as changes in the solvent accessible surface area (Sept and McCammon 2001). According to this work the most favorable pathway in spontaneous nucleation is formation of a longitudinal dimer followed by addition of a third monomer to form a protofilament were both longitudinal and cross-filament interactions are represented. The dissociation constant associated to the formation of the tetramer from this trimer is similar to the dissociation constant of the filament elongation reaction. The formation of the trimer is indeed the limiting step. The energetic cost for association of a monomer to the barbed-end of a trimer or bigger oligomer or filament is lower than the first step of dimerization and trimerization leading to spontaneous filament elongation at the barbed end.

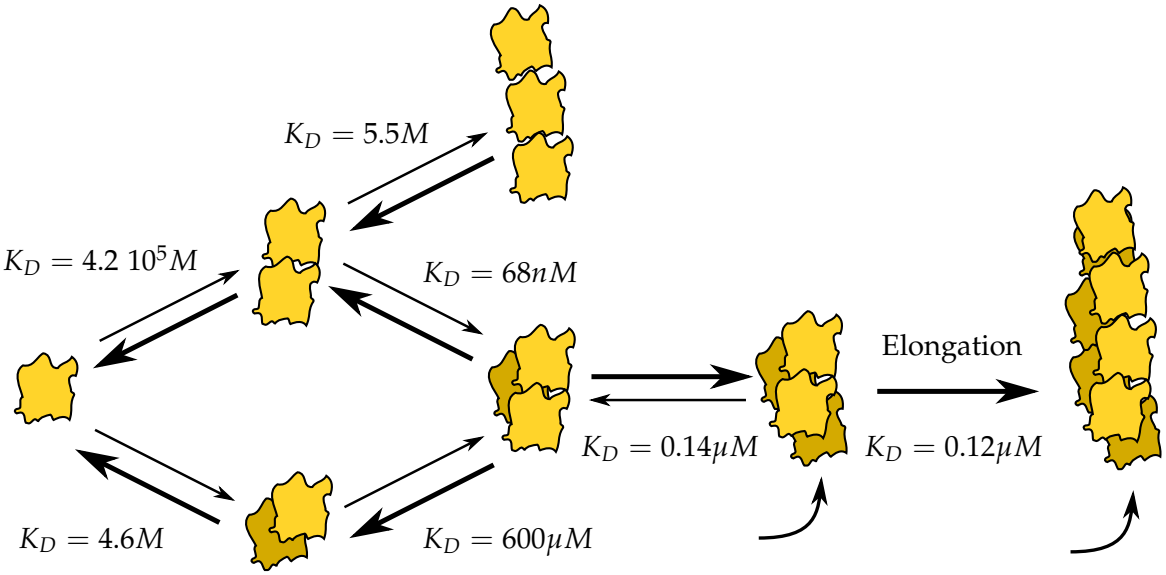


Figure 2.4 – *Thermodynamics of actin nucleation. Dissociation constants were obtained by computer simulation and free energy calculation (Sept and McCammon 2001). The elongation equilibrium dissociation constant is for association of monomers to the barbed-end of filaments.*

In cellulo, nucleation is catalyzed by actin binding proteins which stabilize actin nuclei.

2.3.1.3 Elongation, depolymerization and steady state

Elongation is the addition of actin monomers to a filament. Elongation is polarized : critical concentrations differ for the pointed and barbed end of the filament and for ATP-actin and ADP-actin as shown on figure 2.5 (Pollard and Borisy 2003). Therfore ATP-actin monomers preferably associate with the barbed end of the actin filament.

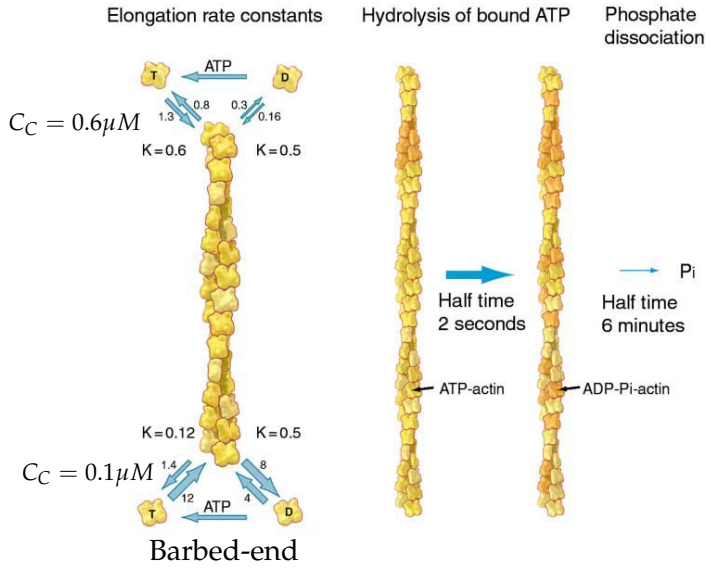


Figure 2.5 – *Thermodynamics of actin elongation, ATP hydrolysis and phosphate release. Dissociation equilibrium constants, association and dissociation rate constants are indicated for pointed and barbed end of the filament and ATP and ADP-actin (Pollard and Borisy 2003).*

Filament assembly is associated with rapid ATP hydrolysis and delayed phosphate release (Korn et al. 1987). ADP-Pi-actin as a similar conformation to ATP-actin. Phosphate release induces a conformational change in actin and finally leads to dissociation from the filament.

At the steady state, ATP-actin monomers associate with the barbed end of actin filaments and are rapidly hydrolysed into ADP-Pi-actin, phophate is released after a while and conformational changes in ADP-actin promote dissociation from the pointed end of the filament. The ADP of dissociated G-actin is then exchanged against ATP to enable reassociation at the barbed end of the filament. *In cellulo*, nucleotide exchange is catalyzed by profilin (Selden et al. 1999) This steady-state is called treadmilling (Disanza et al. 2005).

2.3.2 Structures of the actin filament

The actin filament is a non-convalent helical homopolymer former by the reversible assembly of G-actin molecules. It is 6 to 10 nanometers wide and up to several micrometers long. F-actin can be considered either like a left-handed helix with a 166°rotation per molecule or two intertwined right-handed helices with a 28°rotation per molecule. The filament is stabilized by longitudinal as well as transverse intermolecular actin/actin interactions as shown on figure 2.6. Subdomains 1 and 2 and directed toward the outside of the filament and subdomains 3 and 4 toward the central axis of the filament.

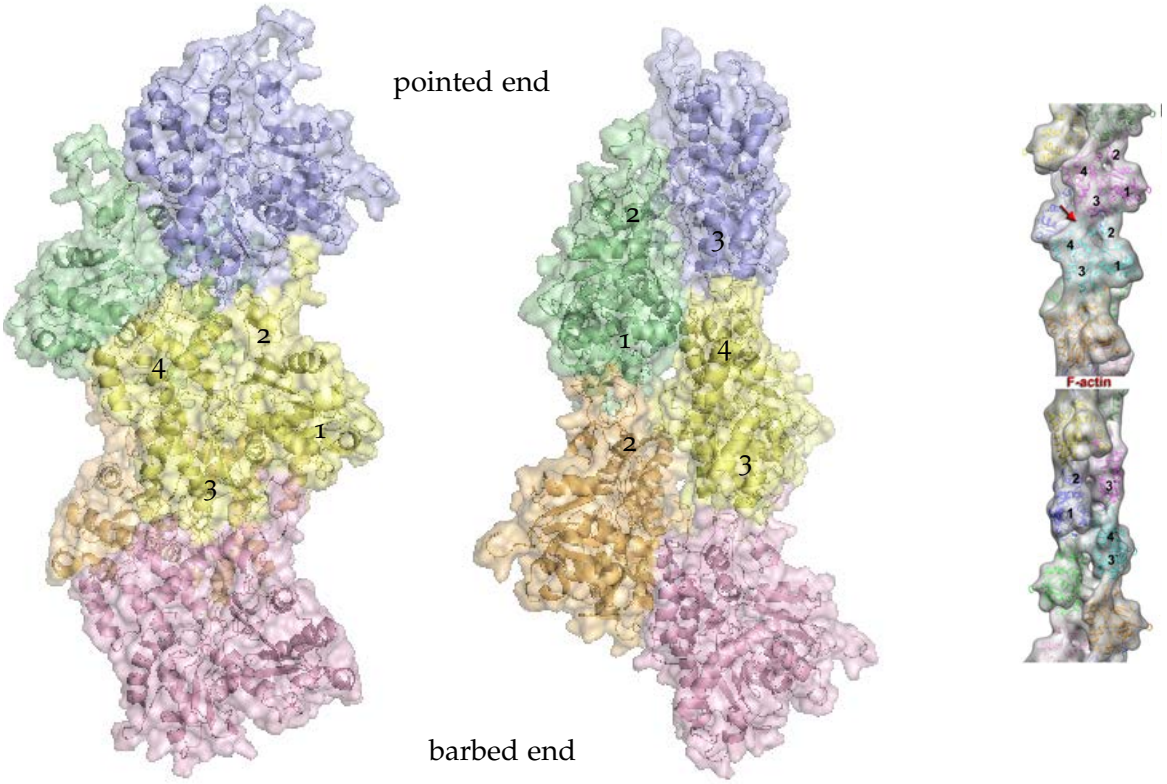


Figure 2.6 – *Organization of the F-actin filament (PDB : 3J8I)*

Structural investigations on actin filaments are hindered by their fibrillar nature which

makes them unfit for crystallization. Besides, actin filaments are highly polymorphic (Galkin et al. 2010) and high resolution structural studies require to isolate one particular rather homogeneous state.

Single actin filaments were observed as early as in 1949 by electron microscopy (Rozsa et al. 1949) but the first molecular model of the filament was proposed after the structure resolution of the actin monomer (Holmes et al. 1990). Oda et al. (2009) further improved this model thanks to high-resolution X-ray fiber diffraction data and were able to show that actin transition from the monomeric to the filamentous state comes with a conformational change in the actin unit. The two major domains rotate by about 20 degrees resulting in flattening of the actin molecule as shown on figure 2.7. More recently near atomic models of canonical and tilted actin filaments were obtained by electron microscopy with a 4.7 angstroms resolution (Galkin et al. 2015). Reduction of actin filaments heterogeneity was achieved by freezing the sample in thin ice.

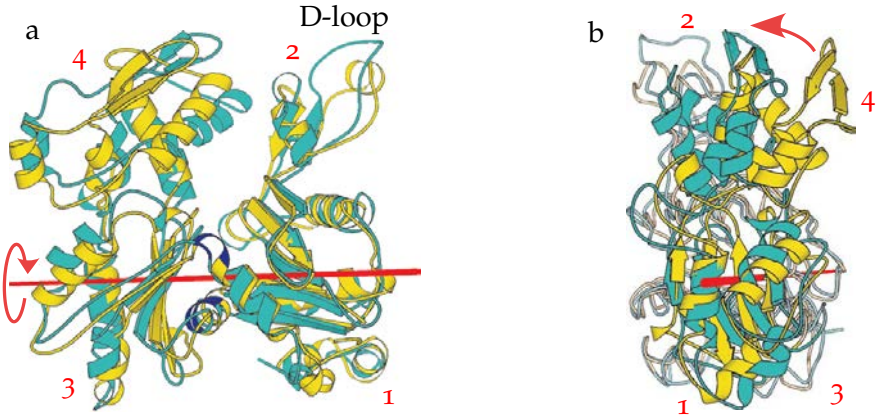


Figure 2.7 – *Transition from the G-actin conformation (in yellow, structure of TMR Ca-ATP G-actin, PDB : 1J6Z) to the flat conformation in F-actin (in cyan, PDB : 2ZWH). Subdomains 3 and 4 are rotated with respect to subdomains 1 and 2 about the rotational axis (red line) in the direction indicated by the red arrow (Oda et al. 2009).*

Longitudinal and cross-filament interactions between actin subunits are highlighted in figure 2.8 and detailed in table 2.2 for filament models deriving from X-ray fiber diffraction (Oda et al. 2009) and electron microscopy (Fujii et al. 2010, Galkin et al. 2015). Although some differences exist between these three models, they all include hydrophobic interactions as well as hydrogen bonds and salt bridges. One should keep in mind that actin filaments are polymorphic and that these interactions are susceptible to be modified by the nucleotide state of actin, mechanical forces applied on the filament or interacting proteins, ions and small molecules.

2.4 Actin cytoskeleton remodeling

Ions, especially cations (Kang et al. 2013), small molecules and proteins impact the formation, length and dynamics of actin filaments. Here we concentrate on the role of actin binding proteins in actin cytoskeleton remodeling.

	Oda et al. (2009)	Fujii et al. (2010)	Galkin et al. (2015)
Longitudinal SD3 ↔ SD4	200-208 ↔ 283-294 241-247 ↔ 283-294	205E ↔ 286D 241E ↔ 324T 244D ↔ 290R 245G ↔ 322P	241E ↔ 324T 244D ↔ 325M 245G ↔ 322P
Longitudinal SD3 ↔ SD2	61-65 ↔ 283-294	40H ↔ 169Y 61K ↔ 167E 62R ↔ 288D 64I ↔ 166Y	38P ↔ 169Y 44M ↔ 168G 61K ↔ 167E 62R ↔ 288D 64I ↔ 166Y
Longitudinal SD1 ↔ SD2	38-49 ↔ 139,140,143 38-49 ↔ 346,351,374 43 ↔ 346 44 ↔ 375	45V ↔ 143Y	44M ↔ 143Y 47M ↔ 352F
Lateral HP plug ↔ SD2	265-271 ↔ 39-42	265S ↔ 40H 268G ↔ 40H 270E ↔ 39R	N/A
Lateral HP plug ↔ SD3	265-271 ↔ 170-174	267I ↔ 173H	267I ↔ 173H
Lateral HP plug ↔ SD4	265-271 ↔ 201-203 265-271 ↔ 285-286	none	271S ↔ 201V/202T
Lateral SD4 ↔ SD1	191-199 ↔ 110-115	191K ↔ 110L 195E ↔ 113K 194T/195E ↔ 110L	194T ↔ 111N 195E ↔ 113K
Lateral SD4 ↔ SD1	none	194T ↔ 177R	none

Table 2.2 – *Intermolecular interactions between F-actin subunits identified in molecular models deriving from X-ray fiber diffraction (Oda et al. 2009) and electron microscopy (Fujii et al. 2010, Galkin et al. 2015). Salt bridges, hydrogen bonds and hydrophobic interactions are highlighted in blue, green and red respectively.*

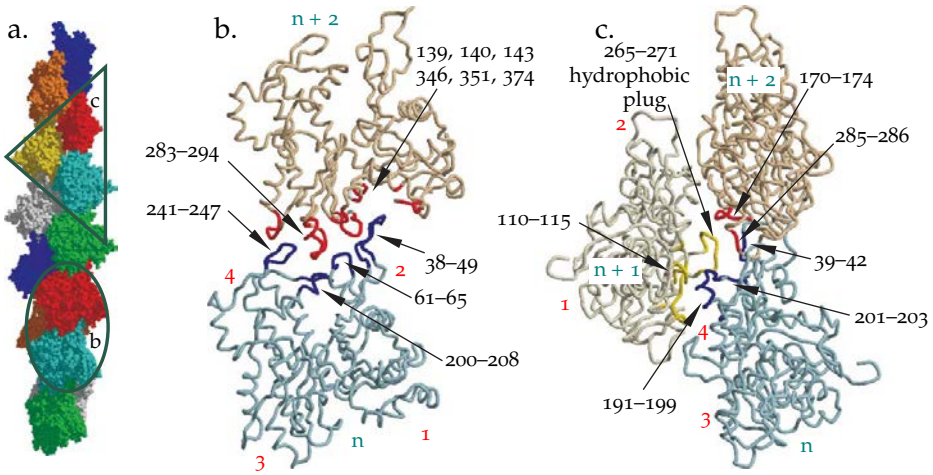


Figure 2.8 – *Intermolecular interactions in F-actin Oda et al. (2009). a. Model of the helical actin filament. b. Longitudinal interactions between subunits n and n+2. c. Cross-filament interactions involving subunits n, n+1 and n+2.*

To accomplish various cellular processes the actin cytoskeleton is carefully remodeled spatially and temporally by a large set of actin binding proteins (ABP) which interact with G and F-actin in a coordinated manner. ABPs functions include monomer sequestration, promotion of filament assembly and elongation, nucleation, filament stabilization, capping, severing and branching, nucleotide exchange and filament anchoring to the cell membrane. Here follow a few examples illustrating some of these functions.

In cellulo, G-actin concentration is in gereral above the critical concentration, providing a pool of available monomeric actin for rapid polymerization. For example, G and F-actin concentration in moving lamellipodia, the actin cytoskeletal network of the leading edge of a motile cell, was estimated at 150 and 500μM (Koestler et al. 2009). Some ABPs, like thymosin β4, sequesters G-actin to maintain the required correct G-actin/F-actin ratio (Xue and Robinson 2013).

Nucleators promote formation new actin filaments either branched to existing filaments for nucleation-promoting factors which activate Arp2/3 or unbranched in the case of formins and WH2 domain-containing nucleators (Campellone and Welch 2010).

Barbed-end filament growth rate is controlled by the interplay of ABPs which bind the barbed-end of the filament to promote elongation (like formins) or to inhibit growth (like capping proteins).

The Arp2/3 complex consists of seven proteins among which the actin-related proteins Arp2 and Arp3 and is involved in filament branching and nucleation. This complex binds a mother filament and the Arp2 and Arp3 proteins constitute the first subunits of a daughter filament which grows with a 70°angle away from the mother filament as shown on figure 2.9.

Actin filaments turnover is driven by depolymerizing and severing proteins. ADF/cofilin severs actin filaments and accelerates the off-rate of ADP-actin monomers at

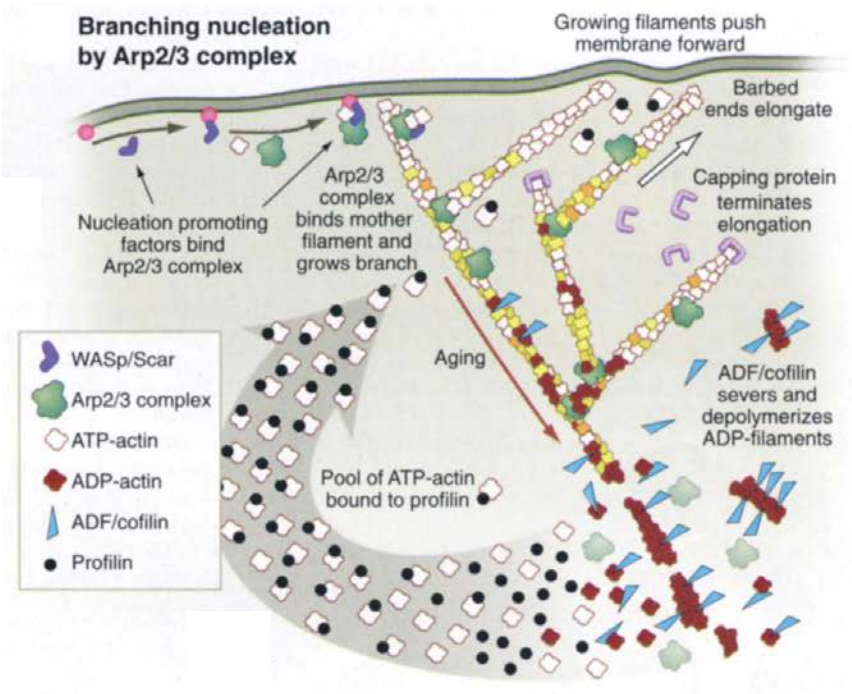


Figure 2.9 – *ABPs that coordinate the formation of branched filaments networks nucleated by Arp2/3 complex (Pollard and Cooper 2009)*

the pointed-end of the filaments. Nucleotide exchange to regenerate a pool of monomeric ATP-actin from depolymerized ADP-actin is catalyzed by proteins such as profilin.

We are interested in the roles of WH2 repeats in actin cytoskeleton remodeling. This particular family of actin binding domains will be presented in the next chapter 3.

2.5 Actin preparation for functional and structural investigations

2.5.1 Actin extraction from cells and tissues

A prerequisite for functional and structural studies is the obtention of high quantities (milligrams) of pure protein. Since actin is very abundant in muscle cells, the most classic method to prepare functional actin for polymerization, interaction assay or structural studies is extraction and purification of α-skeletal actin from rabbit or chicken muscle. Their amino acid sequences are identical to human α-skeletal actin. A protocol was proposed by Spudich and Watt (1971) and improved by Macleanfletcher and Pollard (1980), based on extraction of monomeric actin followed by several polymerization/depolymerization step to separate actin from other endogenous proteins. However this method is very constraining in terms of logistics. It only enables one to prepare wild-type α-skeletal actin and does not allow characterization of the differences between different actin isoforms and understanding the mechanisms of diseases caused by mutations in actin genes such as myopathies (Clarkson et al. 2004).

Other actin isoforms may be purified from non-muscle sources. Adaptation of the protocol is required because actin is much less abundant in these cells and bound to the membrane : 1M tris solutions depolymerize actin filament without denaturation of G-actin and dissociate actin from the membrane (Pinder et al. 1995, Schafer et al. 1998). However, the actin purification yield is rather low : 3mg/200 g of chicken brain, 2 mg/6 L of bovine blood (actin purified from erythrocytes). Besides, this purified actin may be a mixture of different isoforms : mammalian erythrocytes only express β-cytoplasmic actin whereas rat brain cells express β and γ-cytoplasmic actins at a 2:1 ratio (Khaitlina 2001).

2.5.2 Actin overexpression in bacteria and yeast

Recombinant expression in heterologous systems is therefore essential to prepare more diverse or customized actins for which direct extraction from cells is not possible. It also provides increased flexibility and tunability of the accessible actin constructs available. However, actin overexpression is not straightforward. This protein is present in all eucaryotic cells and is involved in many cellular processes. Disruption of the actin cytoskeleton by exogenous actins may be fatal for the host. Another key point is actin proper folding which requires the chaperonin complex containing TCP-1 (CCT).

Expression of recombinant functional actin in *Escherichia coli* is not possible because bacteria lack chaperonin complex CCT and unfolded actin forms inclusion bodies (Frankel et al. 1990). More recently, Tamura et al. (2011) proposed to express actin in *E. coli* at low temperature (16°C) with a cold shock vector (pCold) to facilitate actin folding *in cellulo*. The purification yield after affinity purification on Ni-NTA resin followed by a polymerization/depolymerization cycle was 0.23mg/L of culture which is low for structural studies : 3.75L of culture would be necessary to prepare an NMR sample at 100μM in 200μL.

Actin can be overexpressed in yeast (*Saccharomyces cerevisiae*). However, yeast only has one actin isoform which is essential for the cells viability. Any exogenous actin has to be co-expressed and may fatally disrupt the native yeast cytoskeleton (Sheterline et al. 1995). To overcome this issue, Noguchi et al. (2007) proposed to fuse toxic actin with a sequestering β-thymosin peptide to prevent interaction with yeast cytoskeleton.

2.5.3 Actin overexpression in insect cells

In 2003, Akkari et al. (2003) showed that it is possible to express both wild type and mutant human α-skeletal actin using the baculovirus expression vector system.

Insect cells are widely used for overexpression of viral, eucaryotic and in particular human proteins because they provide chaperones for folding and perform post-translational modifications similar to those in mammalian cells (Drugmand et al. 2012). Besides, they are easier to culture and handle than mammalian cells. The baculovirus expression vector system takes advantage of the fact that polyhedrin which forms the baculovirus envelope and the p10 protein are non-essential in the baculovirus life cycle *in vitro*. Recombinant viruses which contain the gene of interest under one of these two promoters is prepared,

amplified and used for infection of Sf9 cells leading to transient protein expression. The steps for protein expression using the baculovirus/Sf9 system are detailed in the experimental section 4.2.

Purified wild type human α-skeletal actin binds DNase I and polymerizes indicating that it is properly folded. The purification yield was 2mg/L of culture after DNase I affinity chromatography and two polymerization/depolymerization cycles. Since then, various actin constructs were expressed using the baculovirus expression vector system with purification yields up to $10mg/10^9$cells. A non-exhaustive list is presented on table 2.3. Expression of mutants with different polymerizing properties is possible (Joel et al. 2004, Bookwalter and Trybus 2006). The presence of a N-terminal affinity tag to separate overexpressed and endogenous actins generally does not restrains actin overexpression. It was suggested that actin expression using the baculovirus/Sf9 system is promoter dependent (Yates et al. 2007). Early work showed that actin overexpression at the late phase of baculovirus infection reduces the level of polyhedrin (Volkman et al. 1996). Yates et al. (2007) hypothesize that actin is involved in negative regulation of polyhedrin synthesis to explain the different production yield they obtain using a polyhedrin promoter or a p10 promoter. However actin overexpression under the polyhedrin promoter was achieved with good yield by other groups (Ohki et al. 2012).

Applications of recombinant actin expression in insect cells include characterization of the effect of disease causing mutations on actin properties (Bookwalter and Trybus 2006, Miller and Trybus 2008). Engineered non polymerizable actin mutants enabled structural studies previously hindered by actin polymerization such as crystallization of non-polymerizable monomeric actin in the ATP and ADP states (Rould et al. 2006) and study of complexes between actin and tandem WH2 repeats (Ducka et al. 2010, Sitar et al. 2011, Chen et al. 2013).

Reference	vector (promoter)	actin	tag	Yield
Akkari et al. (2003)	BacPak8 (*polh*)	human α-skeletal	aucun	nc
Akkari et al. (2003)	BacPak8 (*polh*)	human α-skeletal	N-ter 6His	nc
Akkari et al. (2003)	BacPak8 (*polh*)	human α-skeletal	C-ter EGFP	nc
Joel et al. (2004)	pAcUW2Bmod (*p10*)	5C drosophila	none	$4mg/10^9c$
Joel et al. (2004)	pAcUW2Bmod (*p10*)	5C drosophila (A204E,P243K)	none	$10mg/10^9c$
Bookwalter and Trybus (2006)	pAcUW51 (*p10*)	human α-cardiac	none	$3mg/10^9c$
Bookwalter and Trybus (2006)	pAcUW51 (*p10*)	human α-cardiac (E99K)	none	$3mg/10^9c$
Rutkevich et al. (2006)	pFastBacHT (*polh*)	human α-cardiac	N-ter 6His	$0.3-0.4mg/10^9c$
Yates et al. (2007)	pFactBacHT-B (*polh*)	human α-cardiac	none	0
Yates et al. (2007)	pFactBacHT-B (*polh*)	human α-cardiac	N-ter 6His	0
Yates et al. (2007)	pAcUW2Bmod (*p10*)	human α-cardiac	none	$3mg/10^9c$
Yates et al. (2007)	pAcUW2Bmod (*p10*)	human α-cardiac	N-ter 6His	$1.5-3mg/10^9c$
Miller and Trybus (2008)	pAcUW51 (*p10*)	human α-cardiac	none	$3mg/10^9c$
Miller and Trybus (2008)	pAcSG2 (*polh*)	human α-cardiac	N ou C-ter FLAG	$0.75mg/10^9c$
Iwasa et al. (2008)	pVL1392-L21 (*pohl*)	human α-cardiac	N-ter Strep	$1mg/2.10^9c$
Ohki et al. (2012)	pFastBac-1 (*polh*)	mouse β	none	$32mg/8*10^8c$

Table 2.3 – *Non-exhautive list of actin overexpression in insect cells for functional and structural studies. nc mean the production yield was not given in the publication. Mutations in the actin sequence are indicated in parentheses. The modified pAcUW2Bmod vector is described in Volkman et al. (1996)*

3 WH2 repeats, disordered and versatile regulators of actin assembly

WH2 domains are intrinsically disordered regions present as single domains or tandem repeats in many actin-binding proteins. Although they bind actin similarly, they exhibit high multifunctionality in the regulation of action assembly. This chapter introduces the general features of this family, briefly presents the basis of interaction with actin and then concentrates on the description of the proteins used in this work : chimeras of thymosin-β4 and of the first domain of Ciboulot, WH2 bidomains of N-WASP and Cordon-bleu.

3.1 General introduction

β-thymosin (βT) and WH2 repeats are small disordered multifunctional domains involved in the regulation of actin cytoskeleton assembly. *In vitro*, βT/WH2 repeats regulate actin assembly by sequestration of actin monomers, promotion of barbed-end actin assembly, nucleation of new filaments, filament severing and filament capping. *In vivo*, βT/WH2 repeats are involved in the remodeling of the actin cytoskeleton during cellular processes such as embryogenesis (Boquet et al. 2000, Carroll et al. 2003), neuronal plasticity, invasion of cancer cells (Nurnberg et al. 2011), or infection by pathogens (Pernier et al. 2013).

β-thymosins, first isolated from calf thymus in the mid-1960's (Huff et al. 2001), form a family of 5kDa peptides which interact with monomeric G-actin and are involved in cellular processes such as cellular migration or endothelial cell attachment and spreading. Thymosins-β4 are ubiquitous and very abundant in lymphocytes and platelets with concentrations up to 500μM and are overexpressed in several tumor (Sribenja et al. 2013).

WH2 domains are actin binding modules of variable length, from 18 to 35 amino acids. WH2 stands for WASP Homology 2 because these domains were first identified in the C-terminal region of the Wiskott Aldrick Syndrome Protein (Derry et al. 1994).

Since then, βT and WH2 domains were identified or predicted in many proteins involved in the regulation of the actin cytoskeleton. They exist as single isolated domain,

like thymosin-β4 which is a single βT domain of 43 amino acids, as tandem repeats, such as tetrathymosin and Ciboulot comprising 4 and 3 WH2 repeats respectively. Single and repeated βT/WH2 domains are also found in larger modular proteins. For example, WASP as one WH2 domain in its C-terminal region and Cordon-bleu has 3 WH2 repeats in its C-terminal region. The SMART database predicts that nearly 900 proteins contain βT/WH2 domains in all species combined (see figure 3.1). Although most WH2 repeats are found in eucaryotes and in particular metazoans, WH2 domain containing proteins were also predicted and characterized in procaryotes and viruses : for example the VopF protein of *Vibrio cholerae* contains three WH2 domains.

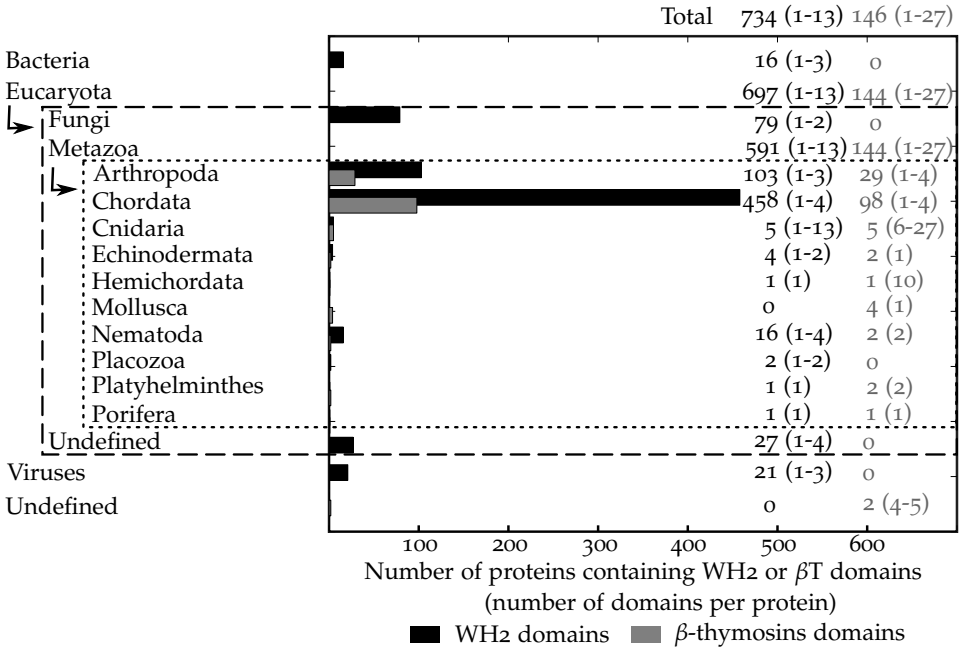


Figure 3.1 – *Taxonomic distribution of proteins with βT (gray labeling) or WH2 (black) domains in the SMART (Simple Modular Architecture Research Tool) database (Letunic et al. 2012). The minimum and maximum number of repeats per protein found within each species is indicated in parenthesis (Renault et al. 2013).*

As for many families of IDPs, βT and WH2 domains have highly variable sequences with about 40 and 20% of sequence similarity within βT and WH2 families respectively (illustrated on figure 3.2) and about 15% of sequence similarities when comparing domains of the two families. Although sequence similarities between the two families are very poor, functional and structural evidence suggest that they belong to a single actin-binding motif family (Edwards 2004, Lappalainen and Mattila 2004).

βT/WH2 domains are mainly identified by a consensus central L++T/V or LKKT/V motif where + are positively charged residues, and a N-terminal extension which can

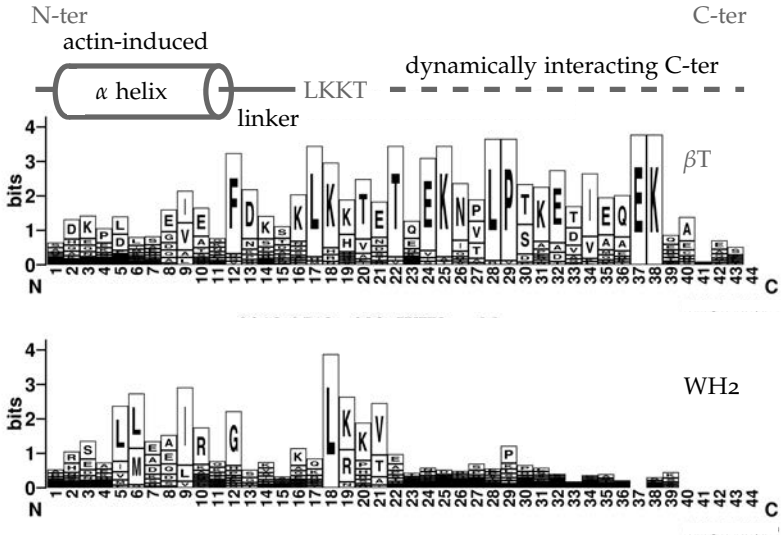


Figure 3.2 – *The sequence similarities of 30 βT and 30 WH2 sequences found in proteins from different species are displayed as a sequence logo for βT (upper) and WH2 (lower) sequence patterns (Crooks et al. 2004). The height of each amino acid letter is proportional to their relative frequency at that position in the sequence. Above are indicated the main sequence elements and how they interact with G-actin(Renault et al. 2013).*

fold into an amphipathic α-helix of 2 to 3 turns that binds the hydrophobic cleft between subdomain 1 and 3 of actin. The highly conserved leucine of the LKKT/V consensus motif interacts with an hydrophobic patch of actin (Ile341 to Ile345) and the positively charged residues probably interact with Asp24 and Asp25 of actin. The rest of the sequence is more variable with C-terminal extensions of variable length and sequences and presence or absence of a linker between the N-terminal amphipathic helix and the LKKT/V motif (Paunola et al. 2002, Chereau et al. 2005, Carlier et al. 2011). β-thymosins share a longer C-terminal extension which forms a second α-helix upon actin binding in most known structures. A sequence alignment of WH2 repeats found in 24 human proteins is presented on figure 3.3.

The structure of thymosin β9 was investigated by NMR spectroscopy in a 60/40 water/fluoroalcohol solution (Stoll et al. 1997). Fluoroalcohol promotes folding into α-helices for regions which already have helical propensities. In this solvent, thymosin β9 forms two α-helices connected by a flexible linker (PDB : 1HJ0).

In absence of organic solvent, βT/WH2 domains lack stable secondary structures. NMR analysis of free thymosin β4 and of the first domain of Ciboulot shows that these protein are mainly unfolded except for the N-terminal α-helix region which retains residual helical structure, especially at low temperatures (Domanski et al. 2004, Hertzog et al. 2004). Overall, there is little quantitative information on the conformational ensembles formed by free intrinsically disordered WH2 repeats. Recently, Elazari-Shalom et al. (2015) proposed a quantitative analysis of the secondary structure propensities in the WH2 domain of WIP by NMR spectroscopy (using chemical shifts analysis, RDCs and PREs). They found that the amphipathic α-helix was partially folded with populations of 1 to 7% for

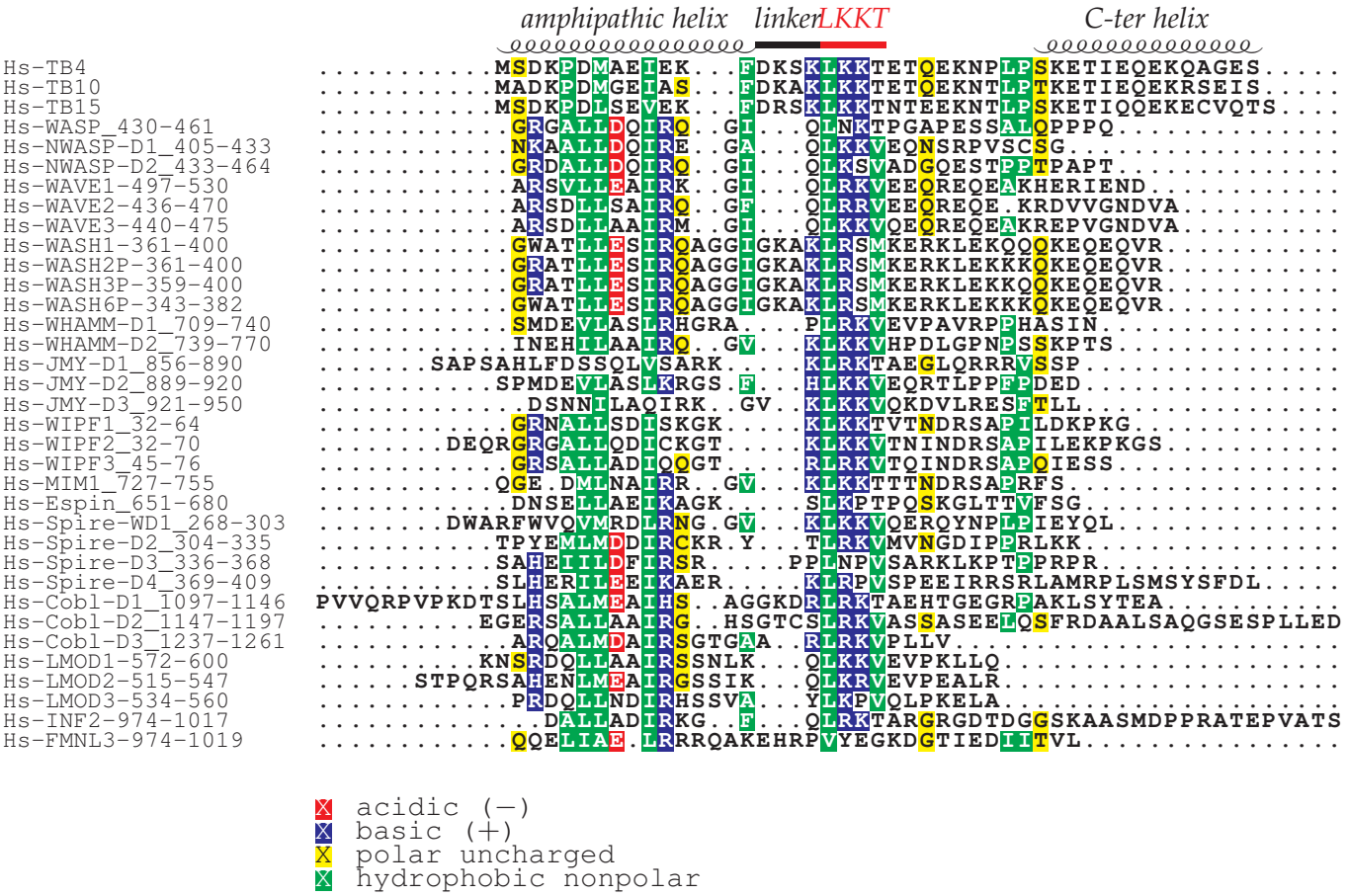


Figure 3.3 – *Sequence alignment of WH2 repeats present in human proteins. D1/2/3/4 designates the first, second third and fourth WH2 domain in the protein. Conserved residues are in green for hydrophobic and in blue for positively charged.*

the folded conformers. This transiently helical region is flanked by two more extended regions with β-sheet populations of 5 to 20% and 5 to 24% before and after the amphipathic helix respectively.

In vitro, single βT/WH2 domains either sequester actin monomers or promote unidirectional actin assembly (profilin-like activity). The sequence-structure/dynamics-function relationship in single βT/WH2 repeats will be developed in section 3.3. WH2 domain repetition increases WH2 repeats multifunctionality and nucleation, fragmentation, filament capping activities were observed *in vitro* for various tandem WH2 repeats.

It should be noted that the domain definition for IDPs is a tricky question since it cannot be defined by a structural unit. In the case of WH2 domains, the amphipathic N-terminal helix and the consensus motif can easily be spotted out but the frontiers of WH2 domains are not so clearly defined. Particular attention should be brought to the choice of domains definitions for functional and structural studies because the function of these domains is defined by a combination of sequence elements which finely tune actin/WH2 interactions along the interface. Thus truncation may have a dramatic effect on the inter-

action with actin. Besides, single or repeated WH2 domains are sometimes surrounded by adjacent disordered conserved regions such as proline-rich regions or lysine-rich regions. These segments affect the *in vitro* function (this point will be discussed later) and should be taken into account to characterize the multifonctionality of WH2 repeats.

3.2 Classification of WH2 domains

WH2 repeats are found in many diverse proteins and are involved in various cellular functions. They can be grouped in different families depending on their mode of action.

Single or repeated isolated βT domains mainly bind ATP G-actin and sequester monomers or promote unidirectional actin assembly (profilin-like activity).

Nucleation promoting factors (NPFs) involved in the activation of the Arp2/3 complex have a C-terminal VCA extension composed of one or several WH2 repeats (V), a connector region (C) and an acidic C-terminus (A) and are predicted to be extensively disordered. They are involved in the formation of branched filament networks in cellular processes involving membrane deformation such as lamellipodia protrusions in migrating cells, phagocytosis, dendritic spine activity, spatial organization of the Golgi or pathogen infections. Taking part in regulation of cell motility, they were also found to be up- and in some cases down-regulated in a variety of cancers (Nurnberg et al. 2011).

These proteins exist in auto-inhibited and activated conformations. In the auto-inhibited form, the VCA extension is buried and not accessible. Upon activation, the VCA is released, the CA region interacts with the Arp2/3 complex and the resulting complex interacts with a preexisting filament. The WH2 repeats in the VCA extension recruit actin monomers to allow nucleation of a new daughter filament (see figure 3.13 for the particular case of WASP/N-WASP proteins). Interestingly some actin nucleation promoting factors only have one WH2 repeat in their VCA extension but N-WASP, WHAMM at JMY comprise two, two and three WH2 repeats respectively. A proposed reason was the improvement of nucleation efficiency by WH2 domains repetition (Yamaguchi et al. 2000, Khanduja and Kuhn 2014) but the underlying structural mechanisms remain elusive.

Although these proteins share common C-terminal features and similar nucleation mechanisms, they have very different N-terminal compositions allowing various modes of regulation and cellular functions (Campellone and Welch 2010).

Activation of these nucleation promoting factors in finely tuned by interacting proteins or lipids and post translational modifications. Factors enabling the conformational switch between the auto-inhibited and active forms of WASP and N-WASP will be detailed in section 3.3.1.

Some WH2 repeats act in synergy with other actin binding like INF2 formin. Formins are another class of actin nucleators which nucleate unbranched actin filaments identified by formin homology domains 1 and 2 (FH1 and FH2). Formins dimerize and bind actin

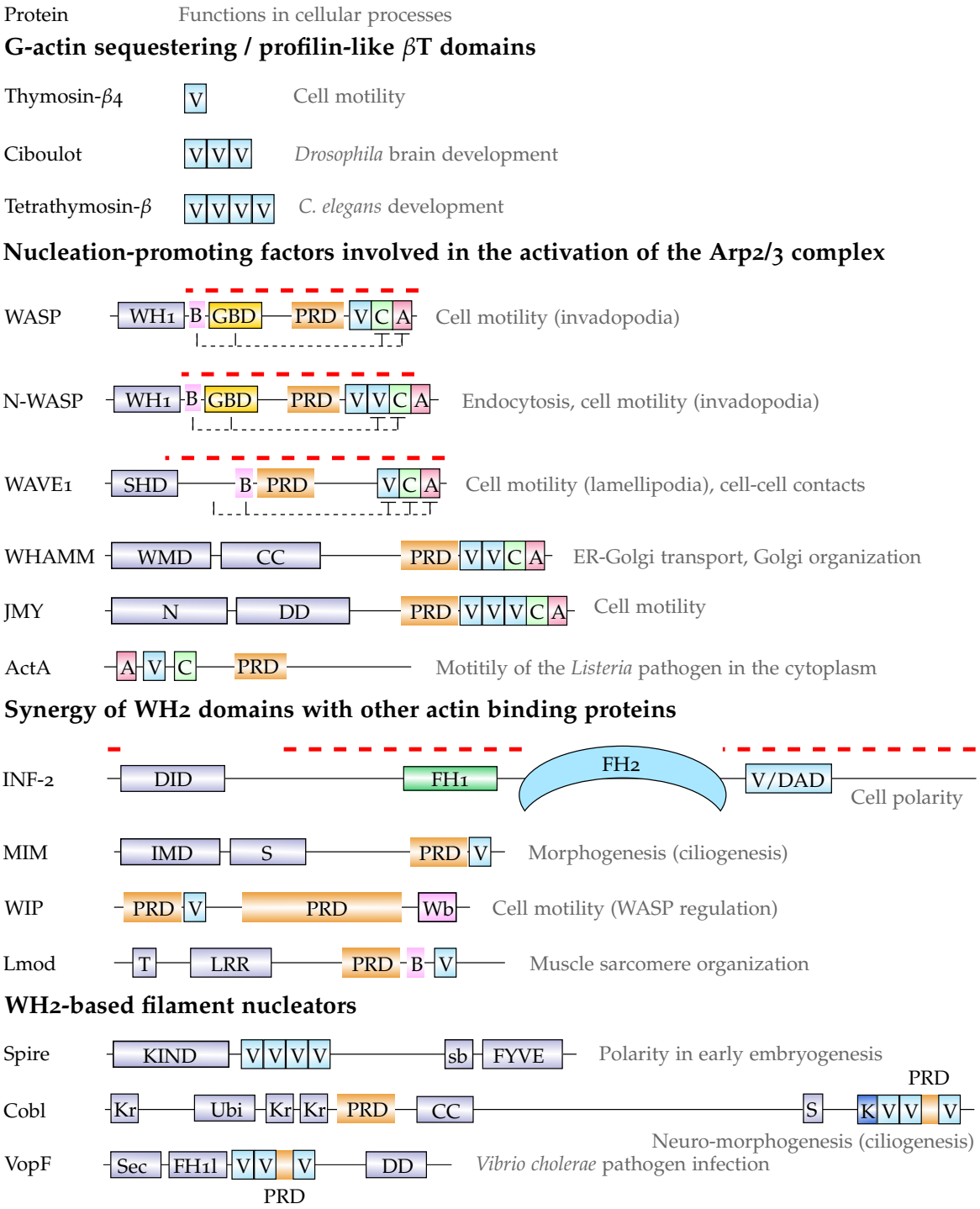


Figure 3.4 – *Example of structural organization and functional diversity (in grey) of proteins containing WH2 repeats (Carlier et al. 2011). WH2 repeats are cyan boxes labeled V. A, acidic domain; B, basic domain, C; connector region; CC, coiled-coil region; DD (spectrin-like), dimerization domain; DID, diaphanous inhibitory domain; FAB, filamentous actin binding; FH1, Formin homology 1 domain; FH1l, FH1-like; FH2, Formin homology 2 domain; GBD, GTPase- binding domain; IMD, IRSp53-MIM domain; KIND, kinase non-catalytic Clobe domain; Kr, 'KRAP' motifs; LRR, leucine rich repeats; mFYVE, modified FYVE zinc-finger; N, N-terminus; PRD, proline-rich domains (orange boxes); S, serine rich region; Sb, spir box; Sec, secretion signal; T, Tropomyosin and actin binding helices; TBR, tubulin-binding region; Ubi, ubiquitin domain; WAHD1, WASH homology domain 1; Wb, WASP binding region; WH1, WASP homology region 1; WHD, WAVE homology domain; WMD, WHAMM membrane interaction domain.*

filaments via their FH2 domain and act as processive filament nucleators as they translocate from elongating filament barbed-ends and prevent barbed-end capping (Chesarone et al. 2010). The rest of their composition is more variable, as for NPFs involved in the activation of the Arp2/3 complex.

INF2 (Chhabra and Higgs 2006) and FMNL3 (Heimsath and Higgs 2012) bind F-actin filaments via their FH2 domain and recruit G-actin monomers via their C-terminal WH2 repeats to accelerate filament elongation.

The last class of WH2 domains containing proteins are WH2-based filament nucleators among which are the Spire, Cordon-bleu, VopF and VopL proteins (Carlier et al. 2011). These proteins are thought to stabilize actin nuclei via interaction with their consecutive WH2 repeats. Besides, WH2 tandem repeats of these proteins are non-equivalent and display high multifunctionality *in vitro* as detailed in the case of cordon bleu in section 3.5.

3.3 Structure and function of single WH2 repeats

Single βT/WH2 repeats may be classified into 3 subcategories. βT domains such as thymosin $\beta 4$ and domains of Ciboulot have a short linker between the N-terminal helix and the L++T/V motif and a long C-terminal region which forms a α-helix in the actin bound form. Other WH2 repeats are referred to as short or long WH2 repeats depending of the length of their C-terminal extension following the L++T/V motif (Chereau et al. 2005).

Single βT/WH2 repeats either sequester actin monomers or promote unidirectional actin assembly at the filament barbed-end (profilin-like activity) *in vitro*. In all cases, filament growth at the pointed-end is inhibited. They preferentially bind ATP G-actin over ADP G-actin (see section 3.5 for an exception) with affinities in the μM range in physiological conditions (Chereau et al. 2005, Didry et al. 2012).

The functional switch between sequestration and profilin-like activities cannot be explained by obvious sequence differences nor by significant changes in affinity for G-actin (De La Cruz et al. 2000, Hertzog et al. 2002, Didry et al. 2012).

Thymosin $\beta 4$ is one of the main actin sequestering proteins in eucaryotic cells. Because of its dynamic nature and instability, the actin:T$\beta 4$ complex is hard to crystallize. Actin binding of thymosin $\beta 4$ was first investigated by NMR (Safer et al. 1997, Domanski et al. 2004). Free T$\beta 4$ is intrinsically disordered in solution as evidenced by the (^{1}H-^{15}N)-HSQC spectrum which shows little dispersion in the proton dimension see figure 3.5. The (^{1}H-^{15}N)-HSQC spectrum of G-actin-bound T$\beta 4$ is dramatically different with important chemical shifts dispersion in the proton dimension. This is in agreement with a global folding upon binding event. Chemical shifts analysis showed that the peptide folded into a N-terminal helix followed by an extended region and a C-terminal helix. These NMR data allowed determination of the fold of the C-terminal region but not its position on the

surface of actin. Irobi et al. (2004) solved the structure of actin bound to a chimera composed of gelsolin domain 1 and the C-terminal half of thymosin $\beta 4$. This method allowed determination of the structure and position of the C-terminal helix bound to actin. A structural model of the complex was proposed from these first results. More recently, Xue et al. (2014) were finally able to solve the structure of fully bound T$\beta 4$ using an actin-T$\beta 4$ chimera produced in yeast. This structure shows both N-terminal and C-terminal helices as shown on figure 3.6 Thus T$\beta 4$ sequesters actin monomers by capping both barded-end and pointed-end of G-actin monomers.

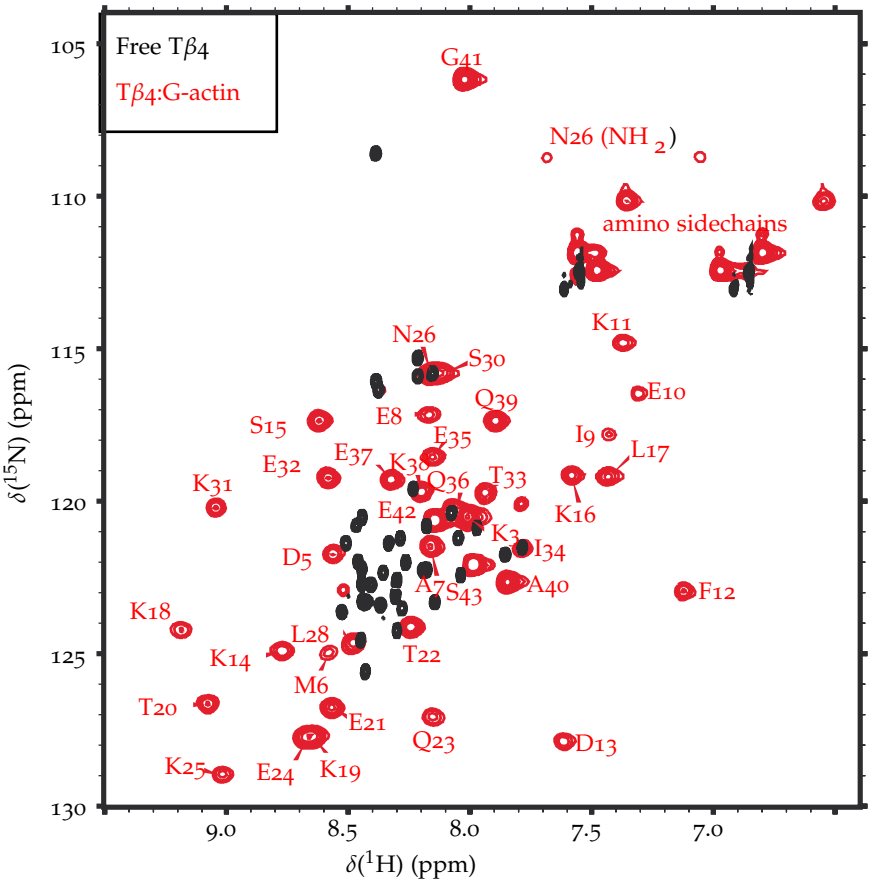


Figure 3.5 – *(^1H-^{15}N)-HSQC spectra of free (in black) and actin-bound (in red) T$\beta 4$ (Domanski et al. 2004).*

Ciboulot regulates actin assembly during brain metamorphosis in *Drosophila* (Boquet et al. 2000). This protein comprises three βT/WH2 repeats. The first domain of Ciboulot, CibD1, has important sequence similarities with T$\beta 4$ but promotes unidirectional assembly *in vitro*.

The same helix-extended-helix fold was observed by NMR for the first domain of Ciboulot as for thymosin $\beta 4$ (Hertzog et al. 2004). The N-terminal amphipathic is longer than that of T$\beta 4$ and forms a three turn helix. The C-terminal helix of CibD1 identified by NMR spectroscopy is not visible on the electron density map obtained by X-ray crystallography (in green on figure 3.7). The electron density is not defined from the third residue after the L++T/V motif until the C-terminus of the protein. Since the differences in function cannot be explained by significant structure differences between these two do-

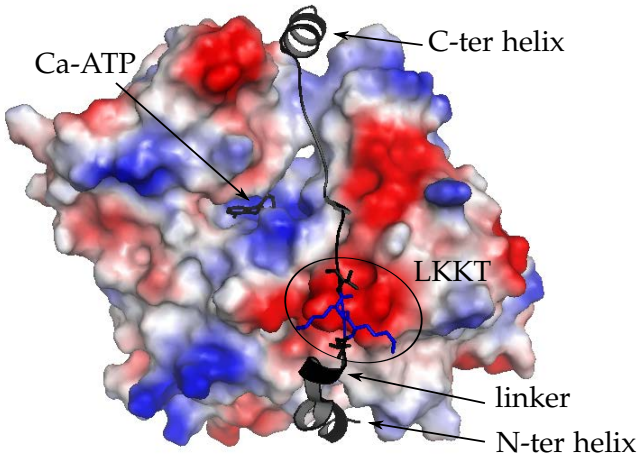


Figure 3.6 – *Structure of the actin:Tβ4 complex obtained thanks to an actin-Tβ4 chimera (PDB : 4PL7). Electrostatic potentials on actin surface are in red for negative/acidic and in blue for positive/basic. The two lysines of the LKKT motif of Tβ4 are shown as blue sticks and their lateral chains are located close to an acidic patch on actin surface.*

mains in the actin bound form, Hertzog et al. (2004) proposed that the functional switch between these two repeats might be explained by different flexibility and dynamics of the actin-bound C-terminal helices. C-terminal region of CibD1 interacts in a more dynamic way with actin and is transiently released, which allows addition of CibD1-bound G-actin monomer to the barbed-end of actin filaments.

Since then, crystal structures were obtained for actin in interaction with various single WH2 repeats : WASP, WAVE2, and WIP (Chereau et al. 2005), MIM (Lee et al. 2007). These crystal structures show that all WH2 repeats adopt a similar fold when they bind actin (see figure 3.7). The N-terminal amphipathic helix binds the hydrophobic cleft between subdomains 1 and 3 of actin. This helix caps barbed-end of actin monomers and thus explains pointed-end filament growth inhibition by WH2 repeats. The positively charged residues of consensus L++T/V motif interact with negatively charged residues on actin surface (Asp24, Asp25, Glu99, Glu100 and Glu334).

Although overall interaction is similar, there are some differences between these WH2 repeats. A conserved arginine at the end of the amphipathic helix of WASP, WAVE and MIM interacts with Thr147 of actin but is not present in the WH2 repeats of WIP and Ciboulot.

The interaction path of CibD1 and Tβ4 is slightly modified between the amphipathic helix and the L++T/V motif because of the residues of the linker which form a 3-10 helix turn (see figures 3.6 and 3.7).

The electron densities are not defined downstream the L++T/V motif for WASP, WAVE and CibD1 WH2 repeats indicating that the central and C-terminal regions of these domains interact with actin in a more dynamic way than their N-terminal regions forming a fuzzy complex or do not interact with actin at all and remain disordered. These two hy-

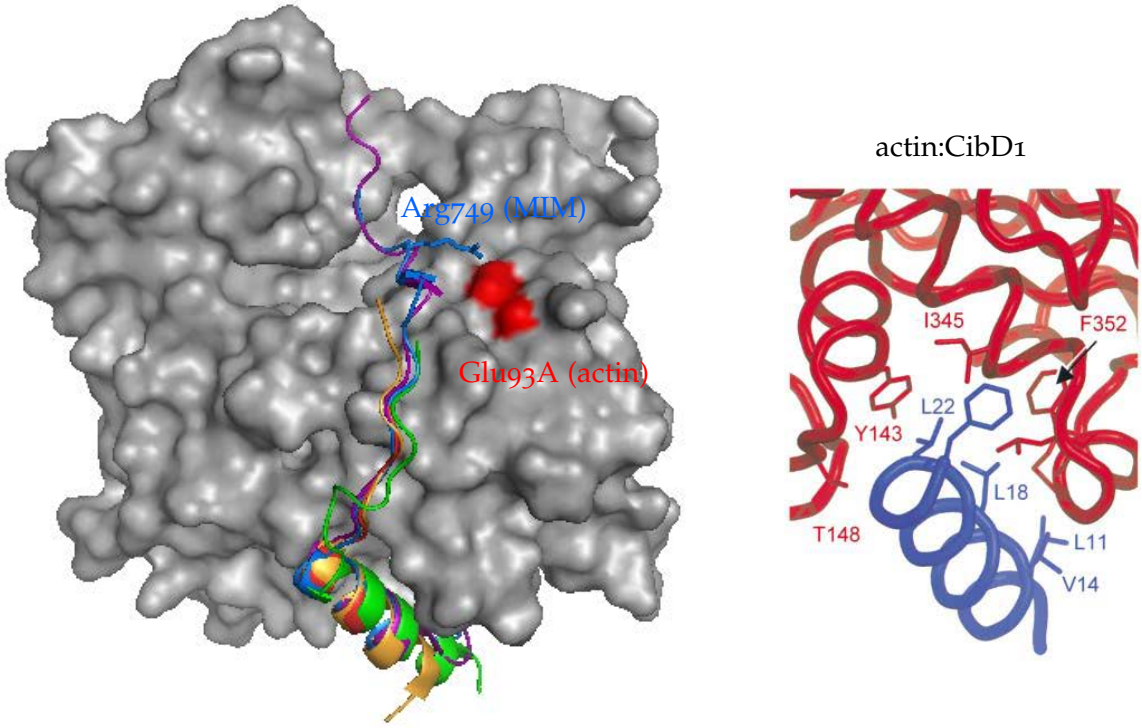


Figure 3.7 – *On the left, superposition of various actin/WH2 complexes crystal structures aligned with the crystal structure of rabbit α-skeletal actin in grey in complex with CiblD1 in green (PDB : 1SQK). WASP WH2 repeat is in red (PDB : 2A3Z), WAVE is in orange (PDB : 2A40), WIP is in purple (PDB : 2A41) and MIM is in blue (PDB : 2D1K). On the right, interactions between the amphipathic helix of CibD1 and actin hydrophobic cleft.*

pothesis are compatible with a profilin-like activity : the pointed-end of actin monomers remain accessible for addition to barbed-end of actin filaments.

Nearly all residues of the central and C-terminal regions of the WH2 repeats of WIP and MIM are visible on the crystal structure. There is a conserved interaction between Arg54 of WIP (Arg749 of MIM) and Glu93 on the surface of subdomain 1 of actin. Once again, despite a very similar interaction determined by crystallography, these two domains have different functions. WIP sequesters actin monomers (Didry et al. 2012) and MIM promotes unidirectional actin assembly (Mattila et al. 2003).

3.3.1 Chimera of thymosin $\beta 4$ and of CibD1

To understand the reasons of the functional switch between sequestering Tβ4 and CibD1 which promotes unidirectional assembly, several mutants and chimera were proposed by the NMR team of the ICSN (*Laboratoire de Biologie et Chimie Structurales*) and analyzed in the team of Marie-France Carlier (*Dynamique du Cytosquelette et assemblage de l'actine*, LEBS). The different constructs and their *in vitro* functions on actin assembly are summarized on figure 3.8 and their affinities for actin in G- and F-buffers in table 3.1 (Hertzog et al. 2004, Didry et al. 2012).

It was proposed that the function of these domains was linked to the dynamics of

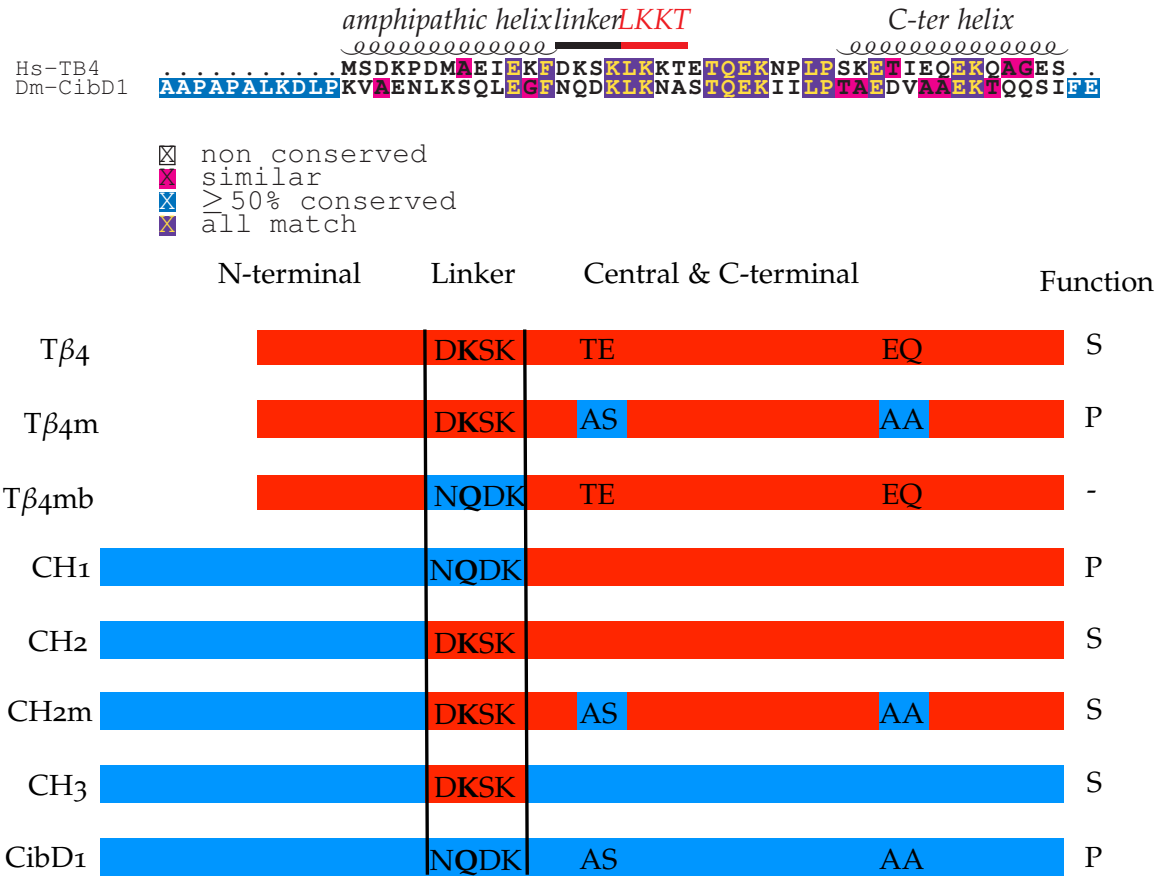


Figure 3.8 – *Schematic representation of the different chimera and mutants of Tβ4 and CibD1. Tβ4 and CibD1 fragments are shown in red and blue respectively. Function of sequestration (S) or promotion of actin assembly (P) is determined in F-buffer at physiological ionic strength.*

the interaction of the C-terminal helix of the βT/WH2 domain and actin. Thus, based on sequence alignments of Tβ4 and CibD1, (Thr20, Glu21) and (Glu35, Gln36) in the central and C-terminal regions of Tβ4 were mutated into (Ala33, Ser34) and (Ala48, Ala49) of CibD1. The obtained mutated Tβ4, Tβ4m promotes unidirectional actin assembly *in vitro*.

However, the nature of these residues is not sufficient to determine the function. CH2 is a chimera comprising the N-terminal helix of CibD1 and the linker, central and C-terminal regions of Tβ4 and sequesters actin monomers *in vitro*. The previously mentioned residues in the central and C-terminal regions were mutated to build similar a CH2m mutant. This mutant sequesters actin monomers, indicating that other sequence elements participate to the modulation of the function.

The role of the linker region was then investigated. CH1, CH2 and CH3 all comprise the N-terminal region of CibD1. CH1 has the linker region of CibD1 and CH2 and CH3 that of Tβ4. CH1 and CH2 have the central and C-terminal regions of Tβ4 and CH3 that of CibD1. CH1 promotes unidirectional actin assembly and CH2 and CH3 sequester actin monomers. Thus replacing the central and C-terminal regions of CibD1 by those of Tβ4 is not sufficient to induce a functional switch. Additionally, chimera which have

Protein	K_D in G-buffer (Ca-ATP-actin)	K_D in F-buffer (Mg-ATP-actin)
Tβ4	0.2 μM	2.3 μM
CH1	0.25 μM	3.5 μM
CH2	0.001 μM	0.5 μM
CH3	0.01 μM	1.54 μM
CibD1	2 μM	8.5 μM

Table 3.1 – *Binding affinities of various Tβ4/CibD1 chimera for actin. K_D values were derived from fluorescence titration curves using actin AEDANS. Addition of CibD1, CH1, CH2 and CH3 did not induce changes in actin fluorescence so titration curves were realized by addition of Tβ4 in presence of competitive amounts of CibD1, CH1, CH2 or CH3.*

the N-terminal region of CibD1 and the linker region of Tβ4 sequester actin monomers regardless of the nature of their central and C-terminal regions. The presence of Tβ4 linker region is also associated with higher actin binding affinities. K_D values in G-buffer increase by 200 and 250 folds from CibD1 to CH3 and from CH1 to CH2 respectively. Therefore the composition of the linker region in also important to determine the function.

Finally, the length of the N-terminal amphipathic α-helix also affects the actin-binding affinity and the function. CH1 and Tβ4mb both contain the linker sequence of CibD1 and the C-terminal region of Tβ4. CH1 contains the long N-terminal α-helix of CibD1 and promotes unidirectional actin assembly. Tβ4mb contains the short N-terminal α-helix of Tβ4 and its affinity for actin is in the millimolar range which is too weak to have an effect on actin assembly.

Therefore, it was not possible to identify a single residue which would determine the function, regardless of the rest of the domain composition. Instead, it appears that several local sequence elements modulate βT/WH2 repeats function.

The structural basis of the functional switch induced by sequence modifications in the linker region was characterized using an integrative structural biology approach and was the PhD work of François-Xavier Cantrelle (Cantrelle 2010).

The structures of actin:chimera complexes were solved by X-ray crystallography. The structures of actin bound CH1, CH2 and CH3 are similar to the X-ray structure of actin-bound CibD1. The N-terminal region forms a three-turn amphipathic α-helix followed by a 3-10-helix turn in the linker region. The electron densities are not defined downstream the LKKT motif, as for actin-bound CibD1. Crystal structures of chimera:actin complexes provide a first insight into the role of the linker region in the stabilization of the interaction between the central and C-terminal regions of chimeras and actin (see figure 3.10). Lys25 of CH3 (Lys14 of Tβ4 linker) forms an intermolecular salt bridge with Glu334 of actin. In CH1, this lysine is replaced by Gln25 (Gln27 of CibD1 linker) and forms a weaker hydrogen bond with Glu334 of actin. Thus the presence or absence of a single salt bridge in the linker region of CH2 and CH1, respectively, is responsible for the functional switch between the two proteins.

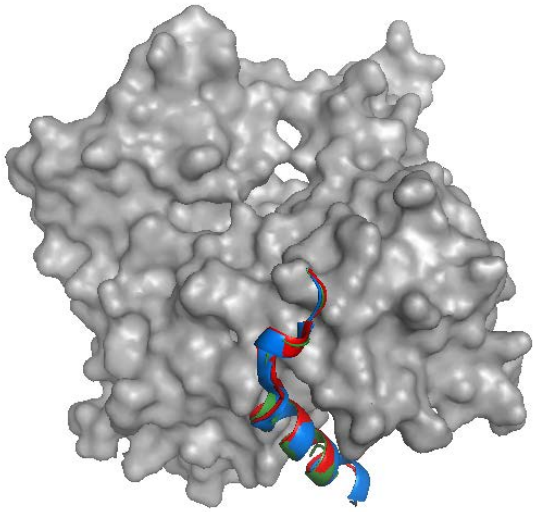


Figure 3.9 – *Superimposed crystal structures of actin:chimera complexes : CH1:actin in red (PBD : 3U8X), CH2:actin in blue (PBD : 3U9D) and CH3:actin in green (PDB : 3U9Z).*

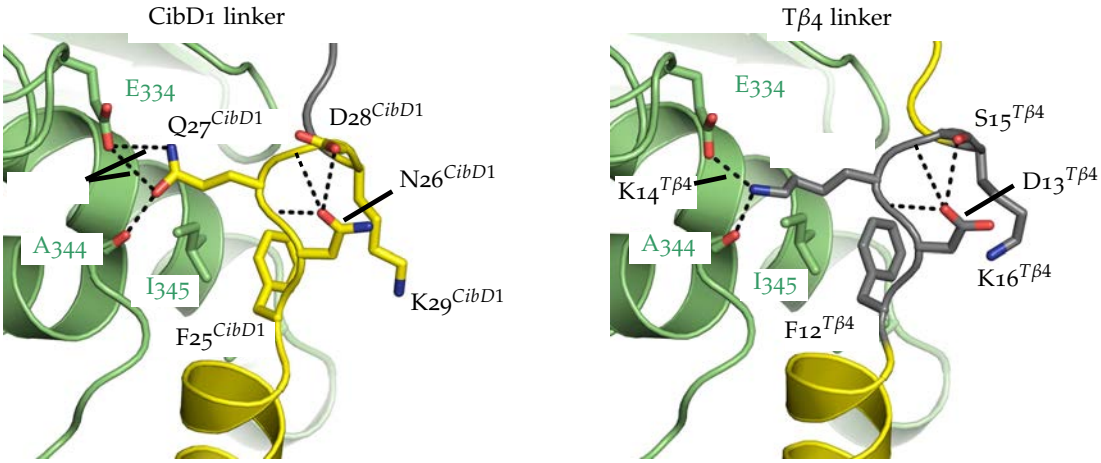


Figure 3.10 – *Interactions provided by CibD1 and Tβ4 linkers in crystal structures of CH1:actin (left, PDB : 3U8X) and CH3:actin (right, PDB : 3U9Z), respectively. Actin, CibD1 and Tβ4 residues are shown in green, yellow and grey, respectively. Side chain electrostatic interactions below 3.8 angstrom are shown. The weak hydrogen-bonds between CibD1 Gln27 and actin Glu334 side chains are substituted in Tβ4 linker by a 2.8 angstrom salt bridge between its Lys14 and actin Gly334 side chains (Didry et al. 2012).*

Chemical shift analysis of the actin-bound chimeras in G-buffer shows that the C-terminal region folds into a α-helix for residues 41 to 54 (C-terminus).

It should be noted that Tβ4, CibD1, CH1, CH2 and CH3 all sequester actin monomers in G-buffer. Variation of function between the constructs is only observed in F-buffer at physiological ionic strength.

The dynamics of the C-terminal regions of CibD1 and CH3 in G- and F-buffer were analyzed by SAXS. The SAXS curves are superimposable for CH3 in G- and F-buffer and for CibD1 in G-buffer and are compatible with a model where the N- and C-terminal regions form α-helices in interaction with barbed- and pointed-ends of actin respectively. This is consistent with the sequestering activity of CH3 in G- and F-buffers. This model fails to describe the SAXS curve for CibD1 in F-buffer, however, SAXS data are compatible with a structural model where one α-helix is folded and interacts with actin and the central region and other helix of CibD1 are disordered and form a flexible tail that does not bind actin. The C-terminal helix is more likely to undergo an order-disorder transition than the amphipathic N-terminal helix which is well visible on electron densities maps of the crystal structures. This model is compatible with a functional switch induced by the increase in ionic strength, from sequestration of actin monomers to promotion of unidirectional actin assembly at filaments barbed-ends.

The dynamics of the C-terminal regions of CH1 and CH2 were analyzed by KCl titration followed by NMR spectroscopy. The free and bound forms of CH1 and CH2 are in slow exchange on the chemical shift timescale. Thus, the loss of actin interaction at residue resolution was monitored by measuring intensity loss in bound peaks or intensity gain in free peaks as the KCl concentration increased. For CH2, variation of peak intensities is similar throughout the sequence (not shown). Figure 3.11 shows that free CH1 peak intensities increase as the KCl concentration increases for the N- and C-terminal regions, indicating that CH1 is partially released as KCl concentration increases in agreement with a weaker binding affinity. However, the C-terminal helix is released first for concentrations as low as 40mM in KCl. This is consistent with the functional switch in CH1 induced by the increase in ionic strength, from sequestration of actin monomers to promotion of unidirectional actin assembly.

Thus SAXS and NMR data provide evidence of release of the C-terminal region in the βT repeats which have a profilin like activity.

3.3.2 Proposed model to explain the functional versatility of βT/WH2 domains

Didry et al. (2012) proposed a model to explain the functional versatility of βT/WH2 domains. The function of βT domains is determined by a combination of parameters including the length of the N-terminal amphipathic helix, electrostatic interactions between actin and the side chains of the linker, central and C-terminal regions of the βT domain and ionic strength.

βT domains interact with actin via a stable N-terminal amphipathic α-helix. This helix

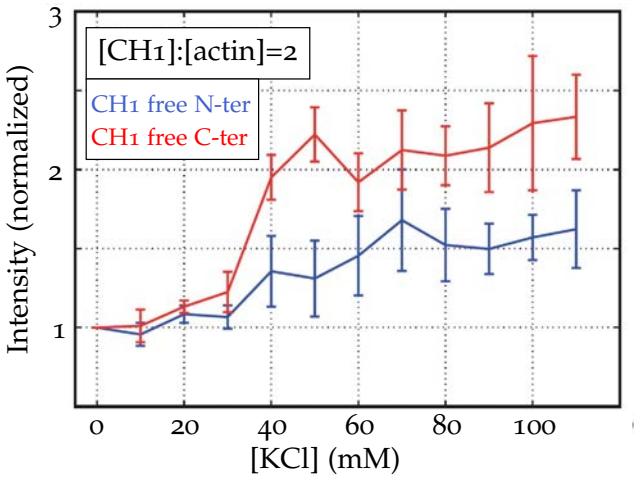


Figure 3.11 – *KCl titration of actin-bound CH1 followed by NMR spectroscopy. The two curves represent averaged intensity values for 5 N-terminal residues in the amphipathic α-helix (Val12, Ala13, Leu16, Glu21 and Phe23) or for 6 C-terminal residues in the C-terminal helix (Glu43, Thr44, Ile45, Glu46, Gln47, Lys49) of CH1 in its free state, as a function of KCl concentration.*

caps monomers barbed-end and prevents filament pointed-end growth. It is followed by a 3-10 helix turn, an extended region and a C-terminal α-helix which interacts with the pointed end of actin monomers. The stability of this C-terminal interaction is determined by the possible electrostatic interactions involving the linker, central and C-terminal regions of the βT domain and actin. At low ionic strength even weak polar interactions and hydrogen bonds are stabilized and the C-terminus of βT domains is bound to actin. As the ionic strength increases, weak electrostatic interactions are screened by ions in solution. In absence of strong electrostatic interactions, the C-terminal α-helix is transiently released and the pointed end of actin monomers becomes available to interact with the barbed-end of actin filaments. In presence of strong electrostatic interactions, charge screening is not sufficient to disrupt interaction between the C-terminal α-helix and actin. This helix remains bound to actin, caps monomers pointed-end and prevents interaction with the filaments barbed-ends hence a monomer sequestration function.

The functional switch between CH1 and CH2 is illustrated on figure 3.12.

βT/WH2 repeats have highly diverse sequences and the presence of the linker is specific to βT domains. However this model may still be transferred to other types of WH2 repeats.

The long WH2 domain of WIP sequesters actin monomers. The side chain of arginine 54 forms an intermolecular salt bridge with the side chain of glutamate 93 of actin. Mutation of this arginine induces a change in the function and the mutant promotes unidirectional actin assembly *in vitro*. Thus Arg54 of WIP plays a similar role as Lys14 of Tβ4 in the stabilization of the C-terminal interaction with actin to ensure a sequestration function.

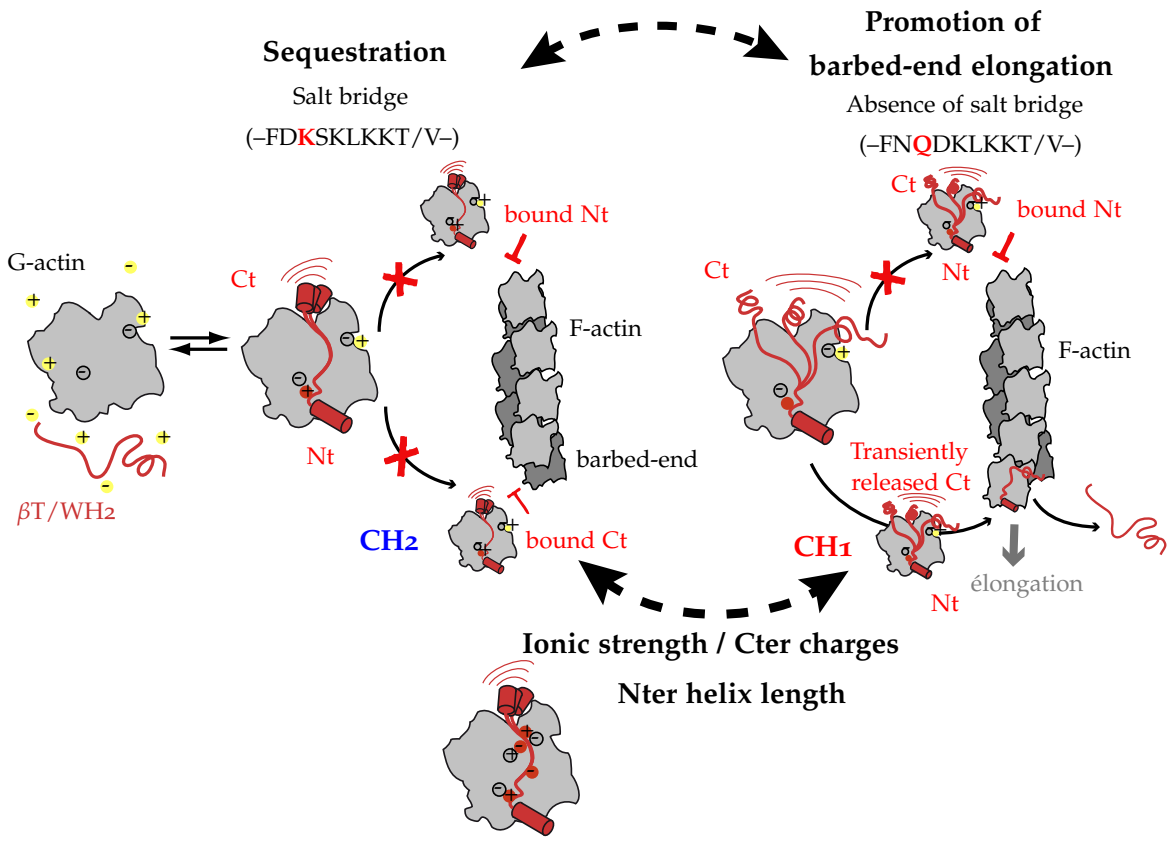


Figure 3.12 – *Proposed model to explain the functional versatility of βT/WH2 domains (Didry et al. 2012). This diagram illustrates the particular functional switch between CH1 and CH2.*

3.4 WASP & N-WASP : NPFs involved in the activation Arp2/3 complex

The WASP and N-WASP proteins are activators of the Arp2/3 complex. WASP stands for Wiskott-Aldrich Syndrome Protein. Mutations in the *WASP* gene lead to the Wiskott-Aldrich Syndrome, a pediatric disorder characterized by eczema, thrombocytopenia, and recurrent infections (Derry et al. 1994). The WASP protein is exclusively expressed in hematopoietic cells. N-WASP (Neural-WASP) is an ubiquitously expressed homolog of WASP particularly abundant in brain cells (Miki et al. 1996). These two proteins share a very similar domain organization except at the C-terminus of the protein : WASP and N-WASP respectively have one single WH2 domain and two WH2 domains in their VCA extensions (see figure 3.4). They are involved in various cellular processes such as cell motility, immune response, vesicle transport along actin filaments (for N-WASP), blood coagulation (for WASP), clathrin-mediated endocytosis (for N-WASP) etc.

3.4.1 Auto-inhibition and activation

Intrinsically disordered regions, shown as red dotted lines on figure 3.4, play an important role in inhibition and activation of the protein. WASP and N-WASP exist in auto-inhibited and active states. Most structural studies presented hereafter were realized on WASP constructs but similar mechanisms are expected for N-WASP because of the same domain organization and sequence similarities between the two proteins.

In the auto inhibited-state, the VCA extension interacts with the central region of the protein : the acidic C-terminal region A interacts with the basic region B situated before the GTPase Binding Domain (GBD) and the connector region C interacts with the GBD. The nature of this interaction was determined by the structure determination of a GBD-C fusion protein by NMR spectroscopy (Kim et al. 2000). The GTPase binding domain is largely unfolded by itself but the GBD-C protein forms a compact assembly comprising a β-hairpin and 4 α-helices in the GBD and a 5th α-helix in the connector region which interacts with the GBD via hydrophobic and polar interactions (see figure 3.13). In this state, Cdc42-GTP and Arp2/3 binding interfaces are both partially occluded on each functional intrinsically disordered GBD and C segments (Kim et al. 2000, Panchal et al. 2003, Hemsath et al. 2005, Kelly et al. 2006). The C/CA interaction with the Arp2/3 complex is too weak to displace these auto-inhibitory interactions (Kelly et al. 2006).

Activation of WASP/N-WASP is triggered by specific and non-specific interactions involving folded and disordered segments in the central B, GBD and PRD regions. For instance, Cdc42-GTPase interaction with the GBD causes a dramatic conformational change involving the β hairpin and the first α-helix (Kim et al. 2000) and disrupts interaction with the connector domain leading to release of the VCA which can interact with the Arp2/3 complex. Cdc42 is not the only activator of WASP/N-WASP and the central region of WASP/N-WASP integrates signals from multiple activators that act in synergy and can be considered as a hub region (Padrick and Rosen 2010). Phopholipid PIP_2 interacts with the basic region B and participates in membrane anchoring (Papayannopoulos et al. 2005). Phosphorylation of WASP/N-WASP Tyr291/Tyr256 mediates interactions with SH2 domains of Src-family kinases which stabilize extended conformations incompatible with the auto-inhibited state. The bacterial effector EspFu of enterohaemorrhagic *E. coli* mimics interactions of the connector domain C with the GBD thus blocking N-WASP in an "active-like" conformation (Cheng et al. 2008).

All these interactions are mapped on WASP and N-WASP sequences on figure 3.14.

3.4.2 WH2 tandem repeat of N-WASP

N-WASP differs from WASP by the presence of two WH2 repeat in its VCA extension. The sequences of these two domains (see figure 3.14) are very similar but the second WH2 domain has more sequence similarity with the WH2 domain of WASP. The VCA extension of N-WASP nucleates actin more efficiently that those of WASP and WAVE,

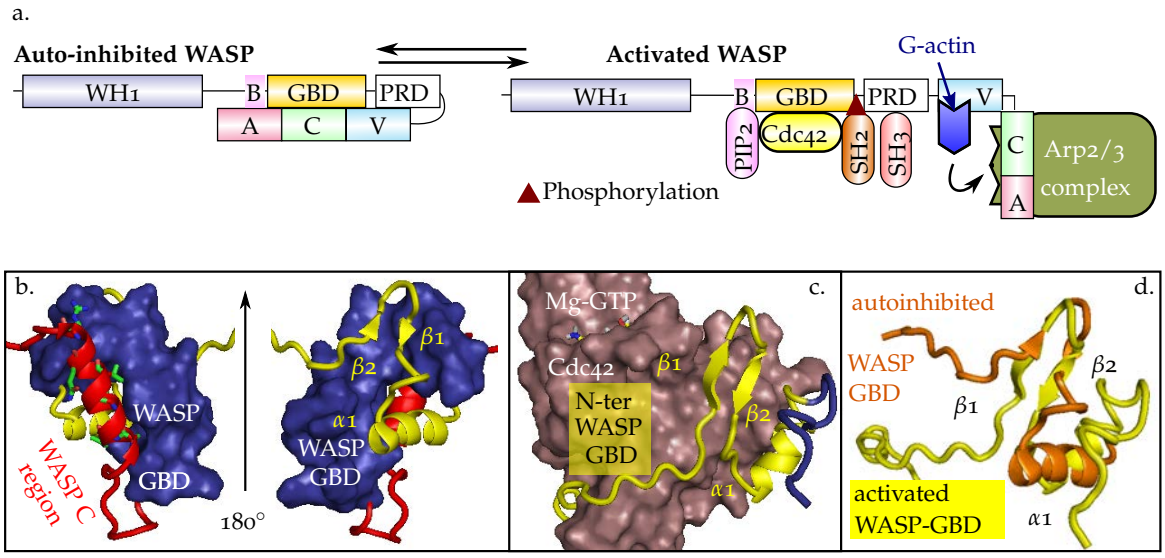


Figure 3.13 – *a. Domain organization and interactions of WASP in its autoinhibited and activated states (Renault et al. 2013). b. In WASP autoinhibited state, its middle GBD (blue ribbon/surface and yellow ribbon) and C-terminal C (red ribbon) IDRs fold with each other to form a small compact folded assembly, which occludes partly Cdc42-GTP and Arp2/3 binding interfaces (Kim et al. 2000). The side chains of C residues involved in hydrophobic and polar interactions at the interface are shown in green (Leu466, Val467, Leu470, Val473, Gln475 and Arg477). c. Activated, extended WASP-GBD (yellow/blue) bound to Cdc42-GTP (brown surface), with WASP-GBD shown in the same orientation as in the right panel of b. (Abdul-Manan et al. 1999). Only part of the C-terminal fragment of the GBD (in blue) is depicted since it unfolds upon binding of the GBD to Cdc42-GTP. d. Superimposition of the two different conformations of WASP GBD bound to C (orange) or Cdc42-GTP (yellow).*

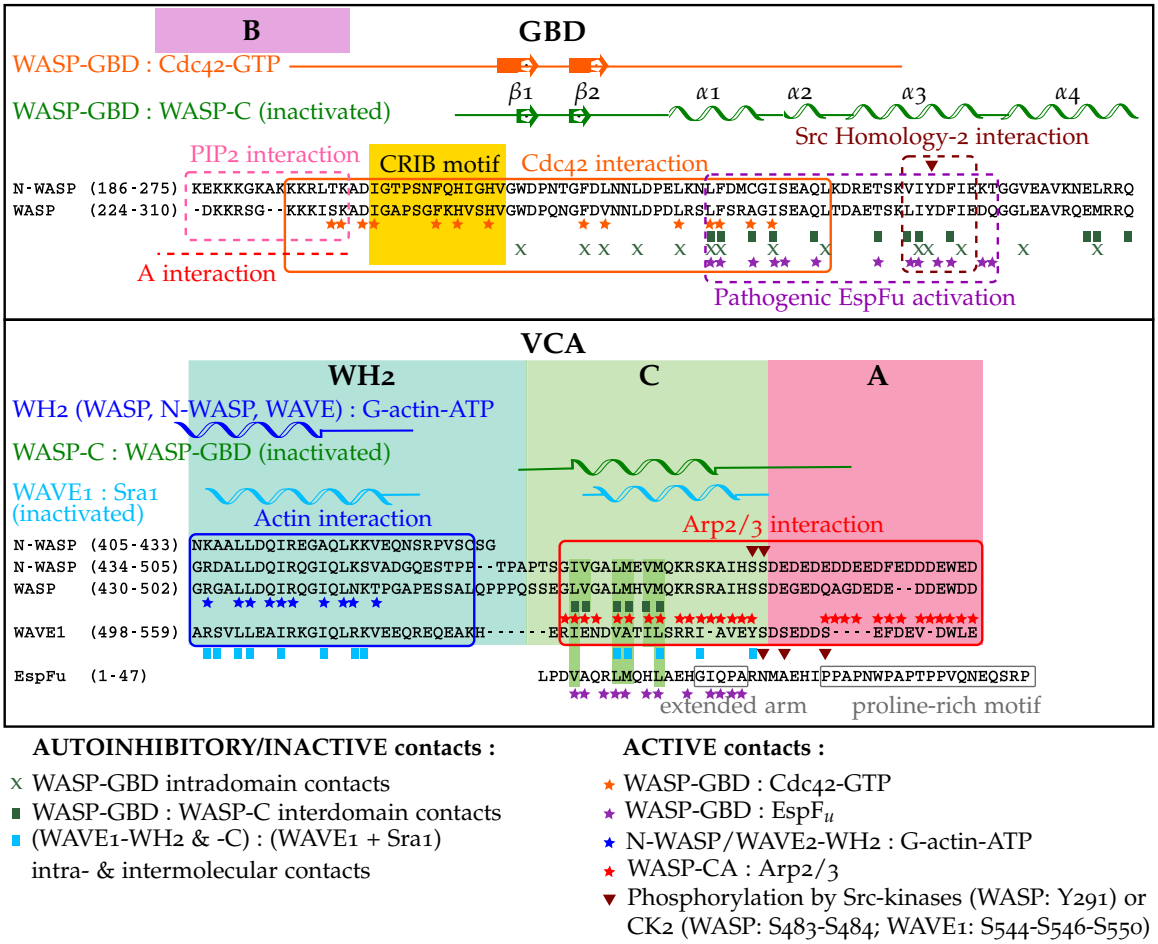
possibly because of the presence of this WH2 tandem repeat. The structure of this WH2 tandem repeat bound to two actins is unknown.

Gaucher et al. (2012) analyzed the interactions of several VCA fragments with actin and Arp2/3 complex. The VC fragment containing the two WH2 repeats and connector region of N-WASP interacts with actin 2 actin monomers with apparent cooperativity as seen in gel-filtration experiments. Two distinct submicromolar binding affinities were measured by fluorescence titration experiments using actin-AEDANS. This time, there was no evidence of cooperativity. Authors conclude that the apparent cooperativity in gel-filtration was due to low values of the dissociation constants determined by fluorescence.

$$V2 \xrightarrow[k_1=0.02\mu M]{+A} V2A \xrightarrow[k_2=0.3\mu M]{+A} V2A_2$$

To avoid actin polymerization, the complex was crystallized in presence of a 2 fold molar excess of VC and the crystal only contained 1:1 VC:actin complexes. The structure of the complex was similar to the WASP:actin:DNaseI ternary complex (Chereau et al. 2005).

Rebowski et al. (2008) engineered a protein containing the two WH2 domains of N-WASP, followed by a duplication of the second WH2 repeat of N-WASP and the C-terminal



AUTOINHIBITORY/INACTIVE contacts :

- x WASP-GBD intradomain contacts
- ▪ WASP-GBD : WASP-C interdomain contacts
- ▪ (WAVE1-WH2 & -C) : (WAVE1 + Sra1) intra- & intermolecular contacts

ACTIVE contacts :

- ★ WASP-GBD : Cdc42-GTP
- ★ WASP-GBD : $EspF_u$
- ★ N-WASP/WAVE2-WH2 : G-actin-ATP
- ★ WASP-CA : Arp2/3
- ▼ Phosphorylation by Src-kinases (WASP: Y291) or CK2 (WASP: S483-S484; WAVE1: S544-S546-S550)

Figure 3.14 – *The B-GBD IDRs of WASP act as a molecular node or hub in the multidomain NPF by integrating many interactions that are mutually exclusive or cooperative (Renault et al. 2013). The amino acid sequences of the B-GBD and V-C-A IDRs from human WASP, N- WASP and WAVE1, or of the N-WASP binding IDR of E. coli EspFu are shown with different labels that underline residues deeply involved in involved in different interdomain and protein–protein interfaces. Secondary structures and interface residues are extracted from the solution or crystal structures obtained in Abdul-Manan et al. (1999), Kim et al. (2000), Chereau et al. (2005), Cheng et al. (2008), Chen et al. (2010), Gaucher et al. (2012). Only residues burying a solvent-accessible surface area > 30 angstrom^2 in the above structures are labeled. Arp2/3 interacting residues of WASP were identified by NMR line broadening experiments (Panchal et al. 2003). The interaction boundaries that are not clearly delimited are shown in dashed lines. The B-GBD IDRs of WASP/N-WASP control 2 interdomain middle-to-tail autoinhibitory interactions, ie with C and possibly A IDRs (GBD:C and B:A), and 4 active interactions, ie with phosphoinositide lipids (B:PIP2), Cdc42-GTP (GBD:Cdc42-GTP), Src family kinases to phosphorylate Y291, and finally Src family SH2 domains after Y291 phosphosphorylation (GBD C-term.:SH2). The inactive and active interactions are mutually exclusive, while active B:PIP2, GBD:Cdc42-GTP and GBD C-term.:SH2 interactions are cooperative. In enterohaemorrhagic E. coli infection, the GBD IDR is targeted by the bacterial effector EspFu (GBD:EspFu) that outcompetes autoinhibited interactions and cellular activators upon a tight binding to GBD, leading to a kind of N-WASP constitutive activation.*

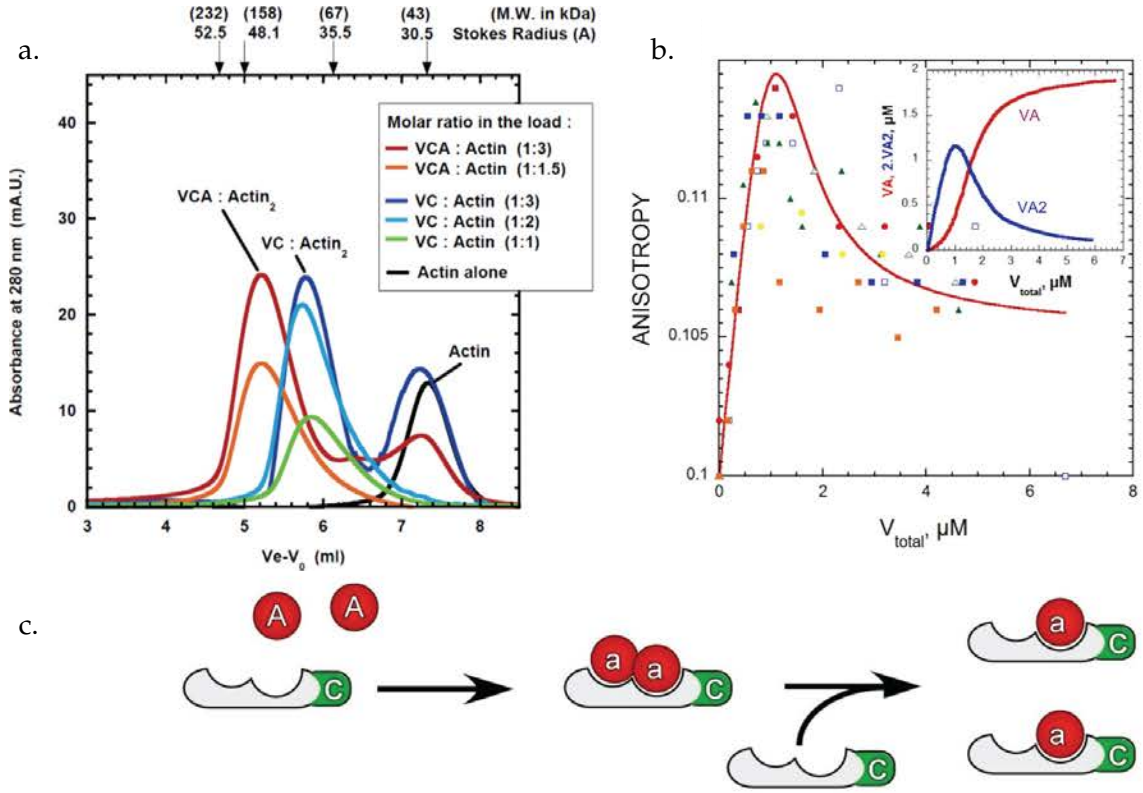


Figure 3.15 – *WH2 tandem repeat of N-WASP interact with two actin monomers (Gaucher et al. 2012). a. Various ratios of actin and VC and VCA fragments of N-WASP were injected on a Superdex200 column. Elution profiles show that actin interacts in an apparent cooperative with VC/VCA to form a 1:2 complex. b. Fluorescence anisotropy of actin-AEDANS as a function of V2 or VC concentration. These experiments are consistent with non-cooperativ binding and two different binding affinities (see below). c. Proposed interaction mechanism : at high actin:V/VC/VCA ratios, actin interacts with the two WH2 repeats of N-WASP. As the concentration of N-WASP fragments increases, 1:1 complexes become predominant.*

helix of Tβ4. This construct binds three actin monomers and Cys374 of the first actin monomer was cross-linked to the first WH2 domain to further stabilize the complex for crystallization.

The crystal structure (Rebowski et al. 2010) shows that the two WH2 repeats of N-WASP interact with actin to form a longitudinal straight actin dimer, different from the twisted longitudinal interactions observed in filaments. However with cross linking and an additional chimeric WH2 repeat, this is rather far from native interactions.

3.5 Cordon-bleu : a multifunctional dynamizer of actin assembly

3.5.1 Organization and function

Cordon-bleu (Cobl) was first identified in mouse embryos and was named after its pattern of expression (figure 3.16) preferentially in the gastrula organizer and in the axial midline (Gasca et al. 1995). This protein is conserved among vertebrates and was shown to be involved in neural tube formation (Carroll et al. 2003), neural morphogenesis and

in particular neuritogenesis and dendritic branching (Ahuja et al. 2007), ciliated sensory hair cells formation (Schueler et al. 2013), regulation of microvilli length in epithelial cells (Wayt and Bretscher 2014) and establishment of left-right asymmetry.

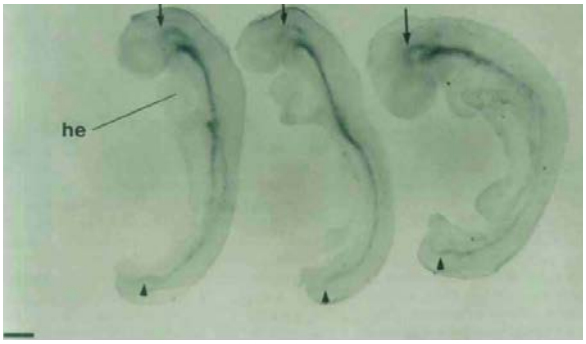


Figure 3.16 – *Expression pattern of the Cobl gene in mouse embryos determined by gene trap experiments (Gasca et al. 1995)*

Vertebrates have a single *Cobl* gene that expresses five isoforms. Residues numbering are those of isoform 1 of human cordon-bleu. Human cordon-bleu is a 1261 amino acid protein (isoform 1) predicted to be mainly disordered (see figure 3.17). It is serine-rich (12% of amino acids) and proline-rich (10% of amino acids). The N-terminal region of the protein comprises an ubiquitin domain and highly conserved KRAP motifs. These motifs are thought to mediate protein-protein interactions but that has not yet been proven (Carlier et al. 2011). The C-terminal region contains three WH2 repeats (A : 1097-1146, B : 1147-1197 and C : 1237-1261) preceded by a lysine-rich domain K. Domains B and C are separated by a proline rich regions.

Cordon-bleu nucleating function is achieved through its C-terminal WH2 repeats (Ahuja et al. 2007).

3.5.2 Multi-functionality of the WH2 repeats of Cobl

In order to understand actin assembly remodeling by WH2 repeats of Cobl, our collaborators at the LEBS analyzed the function of several fragments (Husson et al. 2011). They investigated actin binding by fluorescence and analytical gel-filtration and effect of Cobl fragments on actin assembly by monitoring increase in pyrenyl-actin fluorescence upon actin polymerization.

Results are summarized in table 3.2. Constructs with WH2 domain B at their C-terminus do not include the proline-rich region. CoblABC and CoblKABC were not stable after cleavage of their N-terminal thioredoxine fusion so functional assay were performed on uncleaved proteins.

It appears that WH2 repeats of cordon-bleu are highly multi-functional. Different functions involve G-actin binding (profilin-like and nucleation activities) as well as F-actin binding (severing activity). All constructs have a profilin-like activity and promote unidirectional actin assembly at filaments barbed-ends. The lysine rich domain K is necessary to nucleate actin. This basic sequence enhances transient interactions with acidic actin. The

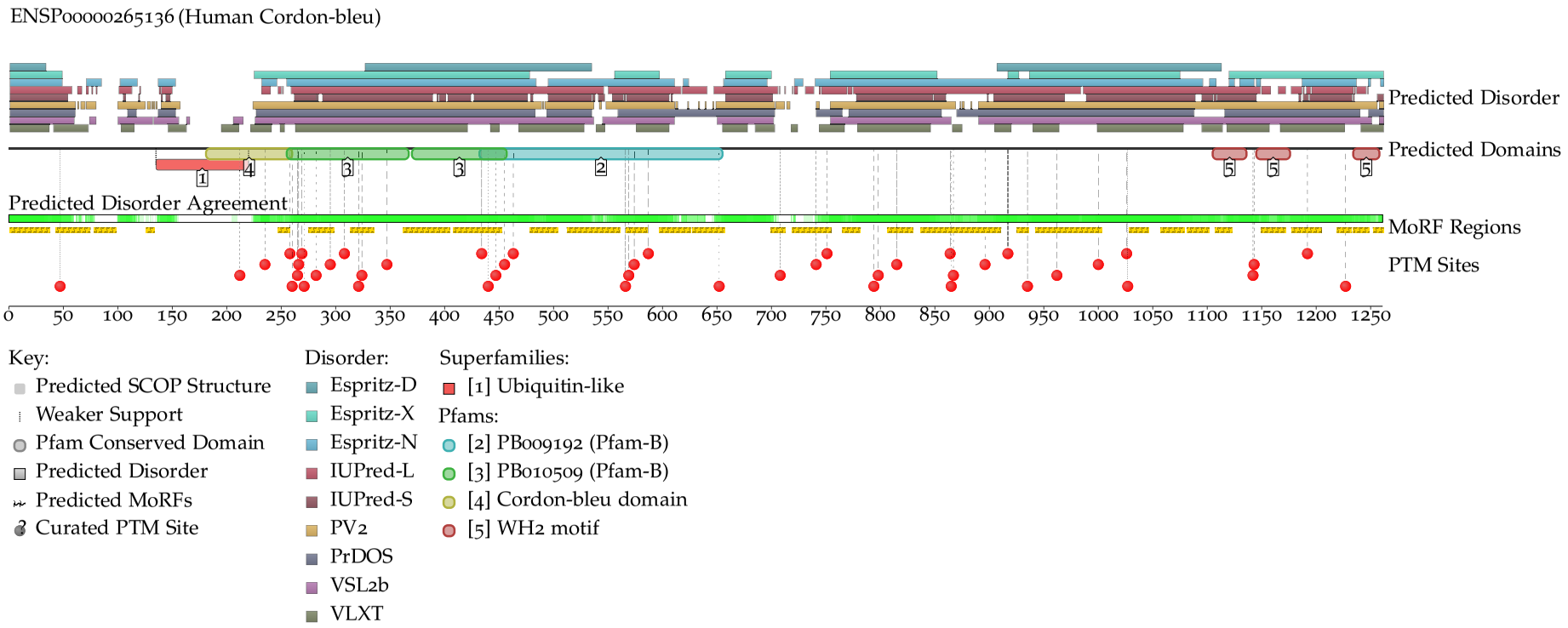


Figure 3.17 – *Organization of human Cobl protein analyzed using the D^2P^2 database (Oates et al. 2013). Disorder is predicted thanks to different algorithms based on various methodologies and predicted disorder agreement is shown in green. SCOP domains are predicted using the SUPERFAMILY predictor. Predicted phosphorylations sites (using PhosphoSitePlus) are shown as red circles.*

Construct	Complex stoechiometry	Profilin-like	Nucleation	Severing
KABC	1:1	+	+	+
KAB	1:1	+	+	+
KA	1:1	+	+	-
A	1:1	+	-	-
B	1:1	+	-	-
AB	1:1	+	-	-
ABC	ND	+	-	-

Table 3.2 – *Multi-functionality of Cobl fragments. Stoechiometry of the complex is determined in the* $10^{-7}M$ *range*

lysine-rich domain K and at least two WH2 repeats are necessary for severing function. Thus KAB is the shortest module to have maximal multi-functionality.

Interestingly, actin filament severing by Cobl is different than severing by Spire. Spire WH2 repeats remain bound to actin after severing thus capping barbed-end of severed filaments (Bosch et al. 2007). On the other hand, Cobl dissociates from severed filaments allowing reannealing.

Another interesting feature is sequestration of ADP G-actin (validated on CoblKAB). This is unusual for WH2 repeat which generally preferentially bind ATP G-actin.

At higher concentrations in the 10^{-5}M range, 1:2 Cobl:actin complexes are formed for AB and KAB. Based on affinity for individual WH2 domains ($K_T = 0.5\mu M$ for A and $K_T = 0.5\mu M$ for B) Husson et al. conclude that in tandem repeats, actin preferentially binds WH2 domain A of Cobl.

Sequence elements responsible for the multifunctional activity of KAB were further analyzed using mutagenesis and domain swapping approaches and only the unmodified KAB fragment displayed severing activity (Jiao et al. 2014).

3.5.3 Structure of a WH2 bidomain of Cobl in interaction with two actin monomers

When I started my PhD, there was no high or low resolution structure showing the interaction of Cobl WH2 repeats with several actin monomers. Actin:Cobl complexes are hard to manipulate at high concentration suitable for structural studies because actin tends to polymerize. Based on analysis of CoblABC sequence, it was proposed that the three WH2 repeats stabilize an actin trimer comprising both longitudinal and cross-filament interaction. Longitudinal intermolecular interactions in the trimer would be stabilized by actin interaction with A and B WH2 repeats of Cobl. The third WH2 domain C, distant from the first two because of the proline-rich region, would stabilize cross filament interactions with a third actin monomer. However, there was no structural information on the interaction of this WH2 repeat ABC with actin to confirm this hypothesis.

Chen et al. (2013) developed a double-mutant strategy to crystallize a complex between mouse CoblAB and two actin monomers. They expressed two non-polymerizable *drosophila* actin mutants in insect cells : AP-actin (Joel et al. 2004) or actinI comprising two substitutions in the subdomain 4 (A204E,P243K) and actinII comprising two substitutions in the subdomain 3 (K291E,P322K). These two actin mutants, expected to form non-polymerizable actin nuclei as shown of figure 3.18 were mixed in equal amounts in presence of CoblAB thus enabling crystallization of the complex.

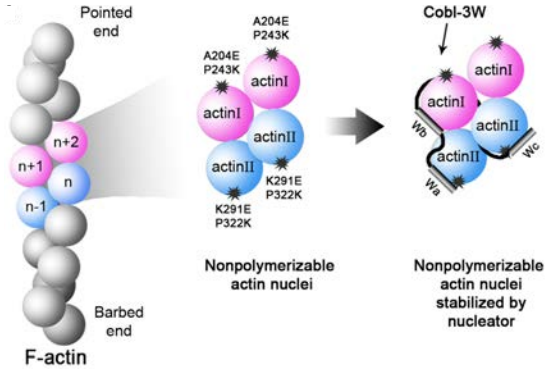


Figure 3.18 – *Double mutant strategy for the crystallization of actin:tandem WH2 complexes (Chen et al. 2013)*

Despite common points with other actin:WH2 complexes, the obtained crystal structure shows unusual features.

The two amphipathic α-helices of domains A and B interact with the hydrophobic clefts of actin monomers 1 and 2 respectively. However, the hydrophilic side of the amphipathic helix of CoblB also interact with the subdomain 4 of actin 1. Intermolecular interactions between Cobl Glu1222, Ser1227 and Arg1234 and actin Glu226, Arg254 and Lys238 repspectively are highlighted on figure 3.19.

The central and C-terminal regions of CoblA interact with subdomains 3 and 4 of actin. Arg1203, Lys1215 and Tyr1218 of Cobl interact with Asp25, Ala331 and Glu214 of actin respectively.

CoblB C-terminal region is defined in the crystal structure and folds into an α-helix which interact close to the nucleotide binding site of actin.

There is a twist between the two actin monomers and the stabilized dimer shows highly different interactions than those observed in the actin filament as shown on figure 3.19.

Intriguingly, the two mutants are not oriented as expected in the crystal : domain A of Cobl interacts with actinI/AP-actin and domain B of Cobl interacts with actinII. So the mutations are not directed toward the outside of the actin dimer.

This structure provides a first insight into CoblAB:actin interaction. However, it would still be interesting to complete these information with solution-state data.

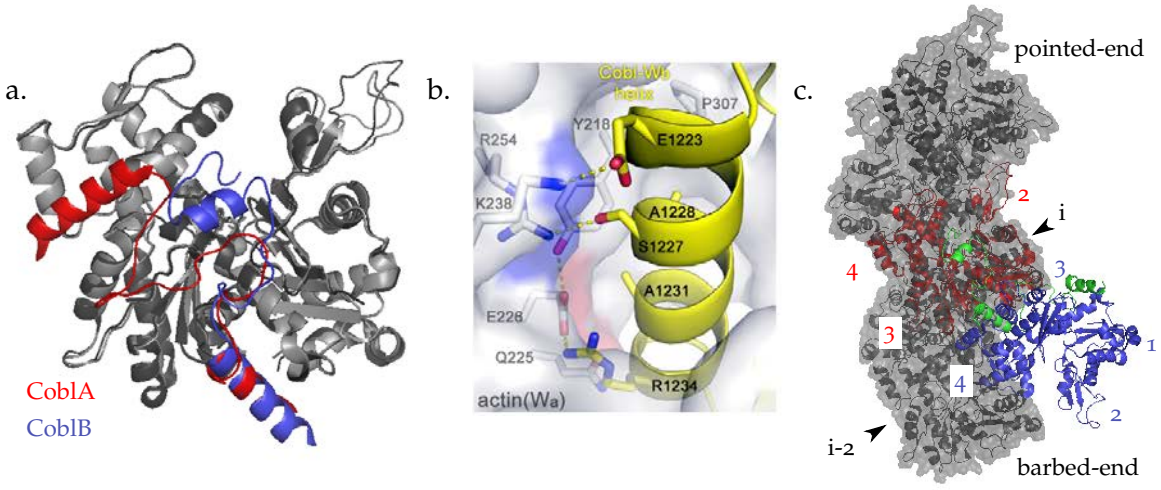


Figure 3.19 – *Features of the CoblAB:actin$_2$ complex (Chen et al. 2013). a. Superimposition of actin-bound Cobl WH2 repeats A and B (PDB : 4JHD). b. Interface between the hydrophilic side of CoblB amphipathic helix and the subdomain 4 of the CoblA-bound actin monomer. c. Comparison of the actin dimer stabilized by CoblAB and the structure of F-actin determined by electron microscopy (PDB : 3J8I, Galkin et al. (2015)).*

3.6 Structural characterization of WH2/actin interactions : challenges and objectives of the present work

This chapter introduced WH2 repeats and a few example of structural studies that provide first insights into the mechanisms by which they regulate actin assembly.

The structural characterization of WH2 repeats and of molecular mechanisms by which they regulate actin cytoskeleton assembly represents a challenge for structural biology for several reasons :

1. Actin polymerizes into filaments at low critical concentrations

Structural investigations are often limited by actin polymerization and filament formation at concentrations suitable for crystallogenesis, NMR or SAXS experiments. It is possible to study native interactions of actin with sequestering proteins. Only a restricted scope of interactions can be characterized this way as many cellular processes are based on dynamic actin cytoskeleton remodeling involving promotion of unidirectional actin assembly, nucleation, fragmentation and so on mediated by the formation of transient complexes with actin binding protein. Thus these intermediate states have to be stabilized artificially. Various strategies include co-crystallization with proteins, such as DNAse I (Chereau et al. 2005) or small molecules, such as latrunculin, which inhibit actin polymerization, cross-linking to stabilize the complex (Rebowski et al. 2008; 2010, Xue et al. 2014) or actin modification to abolish its polymerizing properties by cleavage of the DNAse1 binding loop using subtilisin or ECP32 (Chen et al. 2012), or by point mutations (Rould et al. 2006, Sitar et al. 2011, Chen et al. 2013).

2. WH2:actin complexes are highly flexible

3. WH2:actin nuclei are high molecular weight complexes

WH2-actin interactions cannot be characterized solely by X-ray crystallography of the complexes because some interactions are too dynamic to be defined on electron density maps. One may want to consider stabilization of these dynamic interactions to solve the structure of the full complex but this approach only give a partial view of the picture. Indeed, WH2 repeats adopt a similar fold upon binding actin and dynamics of the complex is crucial to investigate actin assembly regulation mechanisms (Didry et al. 2012). NMR is a method of choice to analyze dynamic protein-protein interactions. But WH2:actin complexes are of high molecular weight for NMR spectroscopy : $\simeq$50kDa for a single WH2 domain in interaction with one actin monomer and $\simeq$100kDa for a tandem WH2 repeat in interaction with two actin monomers. Besides, intermediate binding affinity constants in the μM range may result in intermediate exchange regimes which further complicate the analysis. Finally, since actin cannot be produced in *E. coli*, NMR study rely on observation of labeled WH2 repeats in interaction with unlabeled actin which means looking at the interaction through the small end of the telescope.

Characterization of WH2-actin interaction by NMR spectroscopy therefore requires high magnetic fields, smart pulse sequences and labeling schemes. And it has to be used in combination with X-ray crystallography to map WH2:actin interface and detect possible conformation changes in actin upon interaction with WH2 domains.

The main objectives of this PhD project are the following :

1. Set up production of recombinant actin using the baculovirus expression vector system.
   This will replace constraining purification from rabbit muscle and will allow production of a non-polymerizable actin mutant to facilitate structural studies (Joel et al. 2004). This is especially important to stabilize complexes between G-actin monomers and multifunctional nucleators like WH2 repeats of Cobl. Additionally, isotopic labeling of recombinant actin would open up exciting perspectives for direct analysis of actin dynamics and conformational plasticity by NMR spectroscopy. But methods developments are necessary for isotopic labeling in insect cells

2. Characterize conformational ensembles of free WH2 repeats in absence of actin to identify possible preformed structural elements and see whether their structural and dynamic characteristics can be linked to their functions or structure in the bound form.

3. Characterize interaction between actin and WH2 repeats.
   For single WH2 domains, structure and dynamics of the complexes were extensively studied so we focus on understanding folding upon binding mechanisms and

exchange between the free and bound WH2 domains using appropriate NMR experiments like relaxation dispersion and chemical exchange saturation transfer. For WH2 bidomain, there are little high resolution data on interaction with actin in solution. We focus on the characterization of CoblAB and NWASPV2 binding modes and we want to elucidate the structure and dynamics of these complexes.

The following part presents experimental methods : protein preparation (chapter 4), biochemical characterization (chapter 5) and NMR spectroscopy (chapter 6).

II

Experimental approach

Protein preparation

4

4.1 WH2 repeats

4.1.1 Proteins

WH2 repeats used for this study were overexpressed in bacteria and constructs are detailed in figure 4.1. Dominique Didry (*Dynamique du cytosquelette et assemblage de l'actine*, LEBS) did the constructs of chimera 1 and 2 of thymosin $\beta 4$ and of the first domain of ciboulot and elaborated the expression and purification protocols. These protocols are briefly summarized here, see Didry et al. (2012) for more details. Production and purification of the WH2 repeats of N-WASP and cordon-bleu required optimization of the constructs and of the expression and purification protocols described in this section.

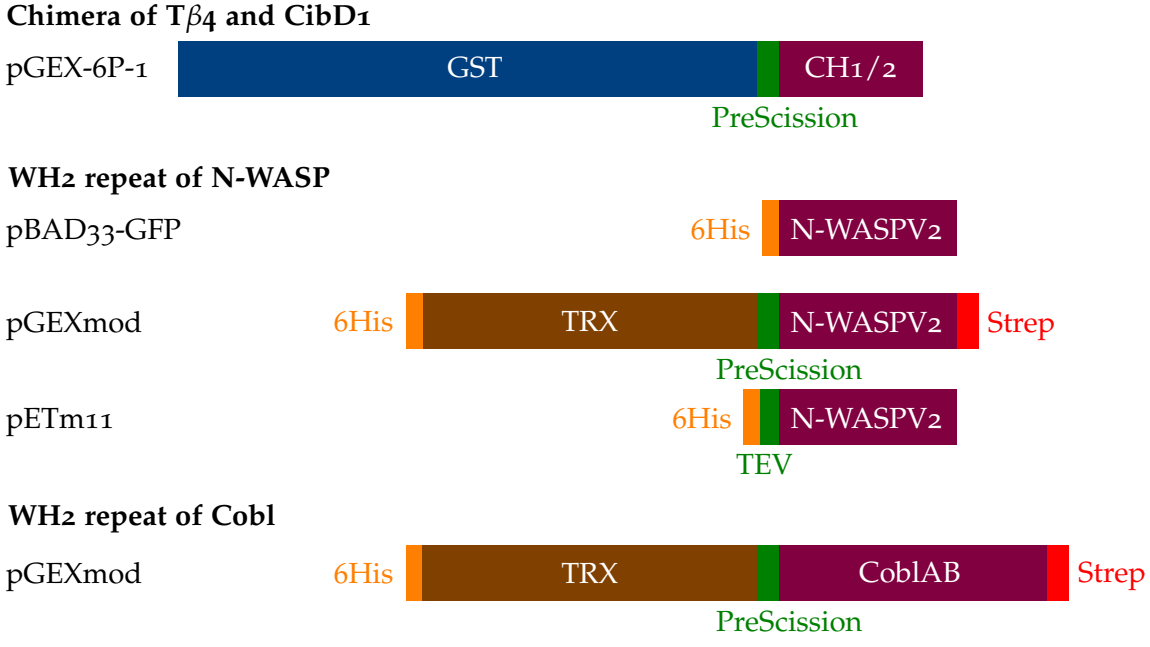


Figure 4.1 – *Constructs for expression of WH2 repeats*

4.1.1.1 Chimera of Tβ4 and CibD1

CH1 contains residues 8 to 29 of drosophila Ciboulot followed by residues 18 to 44 of human thymosin β4. CH2 contains residues 8 to 25 of drosophila Ciboulot followed by residues 14 to 44 of human thymosin β4. Chimera 1 and 2 of Tβ4 and CibD1 were inserted in a pGEX-6P-1 vector using BamHI and XhoI restriction sites. Protein expression was induced by the addition of IPTG to the culture medium. A N-terminal glutathione S-transferase (GST) fusion protein was used for affinity purification. A PreScission cleavage site enabled the removal of the fusion GST.

4.1.1.2 WH2 repeat of cordon-bleu

Constructs of WH2 repeats of cordon-bleu were realized by Clotilde Husson (Husson et al. 2011). Repeats were inserted in a modified pGEX-6P-1 vector (later referred as pGEXmod) using BamHI and XhoI restriction sites. In the pGEXmod vector, the GST fusion protein was replaced by a thioredoxin (TRX) to improve solubility and minimize aggregation during protein expression and to avoid co-purification of truncated fragments due to the GST dimerization. An hexahistidine tag upstream the TRX fusion and a Streptag II at C-terminus position were added for affinity purification. CoblAB contains residues 1159 to 1252 of human cordon-bleu protein.

4.1.1.3 WH2 repeat of N-WASP

Several construct were tested for the expression of the WH2 repeat of N-WASP (later referred as N-WASPV2) as summarized in figure 4.1.

The initial construct was realized by Jean-François Gaucher (Université Paris Descartes) (Gaucher et al. 2012). The WH2 repeat of human N-WASP (residues 392-458) was inserted in a pBAD33-GFP vector under the control of an arabinose-inducible promoter (Guzman et al. 1995). An hexahistidine N-terminal tag enabled affinity purification. Expression in BL21 cells was only transient because of the metabolization of arabinose by the cells. The use of TOP10 cells in which this metabolism is blocked led to a better and more stable expression.

Protein expression in a minimal medium using an arabinose-inducible promoter is tricky because glucose represses expression. Therefore a minimal medium where glycerol was the sole carbon source (Boomershine et al. 2003) had to be used but was incompatible with TOP10 cells because they are leucine and proline auxoroph. The use of BL21 cells might have solved the problem but the metabolization of arabinose might have led to isotopic dilution of the carbon so this construct was set aside.

A synthetic gene coding for N-WASPV2 (residues 402 to 464 of human N-WASP) and optimized for expression in bacteria was ordered from Eurofins MWG and inserted in a

pGEXmod vector using the BamHI and XhoI restriction sites. Expression tests using various *E. coli* strains and induction conditions led to poor protein expression and purification yield so this construct was also set aside.

Finally, the synthetic gene coding for N-WASPV2 was inserted in a pETm11 vector by Véronique Henriot (molecular biology platform, IMAGIF). Protein expression was induced by the addition of IPTG to the culture medium. An N-terminal hexahistidine tag was used for affinity purification and a TEV cleavage site enabled the removal of the tag.

4.1.2 Transformation of competent *E. coli* with vector DNA

E. coli strains used for different vectors are detailed in table 4.1. The human DNA sequence coding for CoblAB contains 5 rare codons (3 coding for arginine and 2 coding for leucine). Therefore, this construct was transformed into Rosetta *E. coli* cells.

Vector DNA		*E. coli* strain	
ID	antibiotic	ID	antibiotic
pGEX-6P1(CH1)	ampicillin	BL21	none
pGEX-6P1(CH2)	ampicillin	BL21	none
pETm11(NWASPV2)	kanamycin	BL21 Star (DE3)	none
pGEX-6P1mod(CoblAB)	ampicillin	Rosetta	chloramphenicol

Table 4.1 – *E. coli strains used for protein production*

Electrocompetent *E. coli* cells were thawed on ice. 1 to 2μL of plasmid DNA was added to the cells, the cells were mixed gently, transferred into an electroporation cuvette and electroporated at 2500V. Cells were immediately transferred into 1mL of LB medium without antibiotics and incubated 1 hour at 37°C and 180rpm. Cells were pelleted and the pellet was resuspended in 100μL of LB medium. 20 and 80μL were plated on LB agar plates containing antibiotic. Plates were incubated overnight at 37°C. Plates containing isolated colonies were stored at 4°C for a maximum period of a month.

4.1.3 Protein expression

4.1.3.1 Production of unlabeled protein

Detailed conditions for expression of various proteins are given in table 4.2.

Generally, 20mL of LB medium with antibiotic were inoculated with a single colony and incubated overnight at 37°C and 180rpm. The overnight culture was added to 1L of LB medium without antibiotics so that the initial OD at 600nm (OD_{600}) ranged between 0.02 and 0.05. The culture was incubated at 37°C until the OD_{600} was between 0.6-0.8.

Overexpression of unlabeled protein was induced by addition of IPTG or arabinose.

The culture was further incubated for 2 to 4 hours. Cells were pelleted at 3000g, 4°C for 10 minutes. The pellet was washed with PBS and cell were pelleted at 6000g, 4°C for 20 minutes. The pellet was frozen in liquid nitrogen and stored at -80°C.

4.1.3.2 Uniform ^{15}N, ^{13}C-labeling

For production of ^{15}N- or ^{15}N-^{13}C-labeled proteins, an additional step was required to transfer the cells to a minimal M9 medium containing ^{15}N-ammonium chloride and ^{13}C-glucose as sole nitrogen and carbon sources (Marley et al. 2001). See appendix for the detailed composition of M9 medium.

The preculture and culture steps were carried out in LB medium as described in the previous paragraph. Cells were then pelleted at 3000g, 4°C for 10 minutes. The pellet was resuspended in 1L of ^{15}N- or ^{15}N-^{13}C-labeled M9 medium. The culture was incubated for 1 hour at 37°C and 180rpm to get rid of the remaining unlabeled nutrients. Overexpression of labeled protein was induced by addition of IPTG. The culture was further incubated for 4 to 6 hours. Cells were pelleted at 3000g, 4°C for 10 minutes. The pellet was washed with PBS and cell were pelleted at 6000g, 4°C for 20 minutes. The pellet was frozen in liquid nitrogen and stored at -80°C.

Protein	Construct	Induction	Duration	Temperature
CH1	pGEX-6P1	1mM IPTG	3-4h	37°C
CH2	pGEX-6P1	1mM IPTG	3-4h	37°C
NWASPV2	pETm11	1mM IPTG (LB) 2.5mM IPTG (M9)	1.5h (LB) 6h (M9)	37°C
CoblAB	pGEX-6P1mod	1mM IPTG (LB) 2mM IPTG (M9)	2.5h (LB) 6h (M9)	37°C

Table 4.2 – *Protein overexpression conditions*

4.1.3.3 Uniform ^{2}H-^{15}N-^{13}C-labeling

Direct observation of the interaction between tandem WH2 repeats and several actin monomers by NMR is challenging because of the size of the complex (MW>100kDa) which results in fast relaxation rates and peaks broadening.

Uniform random incorporation of deuterium in the protein backbone reduces relaxation due to ^{1}H-^{1}H relaxation and ^{1}H-^{1}H scalar couplings leading to significant sensitivity gains. *E. coli* tolerates replacement of H_2O by D_2O but an adaptation step is required and bacterial growth is slower. There are various patterns for deuterium incorporation. Perdeuteration aims at replacement of all non-exchangeable protons by deuterons and is achieved by overexpression in 100% D_2O and use of deuterated glucose. Random fractional deuteration introduces deuterium at random positions and is achieved by overexpression in various H_2O/D_2O ratios or use of protonated glucose. Here CoblAB was

overexpressed in M9 medium containing 100% D_2O, $^{15}NH_4Cl$ and protonated ^{13}-glucose. Deuterium incorporation is expected to be around 80%.

20mL of LB medium with antibiotic were inoculated with a single colony and incubated overnight at 37°C and 180rpm. The overnight culture was added to 500mL of LB medium with antibiotics so that the initial OD at 600nmOD_{600} ranged between 0.02 and 0.05. The culture was incubated at 37°C until the OD_{600} was 0.8. Cells were pelleted at 3000g, 4°C for 10 minutes. The pellet was resuspended in 1L of unlabeled M9 medium ($OD_{600} = 0.4$). The culture was incubated at 37°C and 180rpm until the OD_{600} was 0.8. Cells were pelleted at 3000g, 4°C for 10 minutes. The pellet was resuspended in 100mL of ^{2}H-^{15}N-^{13}C-labeled M9 medium. The culture was incubated for 1 hour at 37°C and 180rpm to get rid of the remaining unlabeled nutrients. 900mL of ^{2}H-^{15}N-^{13}C M9 medium were added and the culture was incubated at 37°C and 180rpm until bacterial started to grow again ($OD_{600} \simeq 0.9$). Overexpression of labeled protein was induced by addition of IPTG. The culture was further incubated for 6 hours. Cells were pelleted at 3000g, 4°C for 10 minutes. The pellet was washed with PBS and cell were pelleted at 6000g, 4°C for 20 minutes. The pellet was frozen in liquid nitrogen and stored at -80°C.

4.1.3.4 Optimization of the expression protocol for CoblAB

Initially protein expression was induced with 1mM IPTG at 16°C for 24h to avoid protein aggregation (Husson et al. 2011). Using these conditions, only a fraction of the expressed protein bound the Histrap column and the eluted fraction contained truncated forms of CoblAB. Increase control of protease inhibition during the first steps of purification did not improve binding to the column nor decrease the fraction of truncated forms. Therefore we decided to optimize the expression conditions to reduce *in cellulo* proteolysis. The screened conditions for expression in LB medium are detailed in table 4.3. Protein expression was assessed by SDS-PAGE analysis of bacterial extracts (see figure 4.2).

Temperature	IPTG concentration	Duration of expression
37°C	1mM IPTG	1 ; 2 & 4h
25°C	2.5mM IPTG	1 ; 2 ; 4 & 6
16°C	1mM IPTG	2 ; 4 ; 6 & 24h
16°C	2.5mM IPTG	1 ; 2 ; 4 ; 6 & 24h

Table 4.3 – *Screening conditions for CoblAB overexpression*

For conditions showing significant protein expression, the pellets of 5mL of culture were lysed using lysozyme. The lysates were clarified by centrifugation. The supernatants were incubated with Ni beads and the beads were washed with a 50mM imidazole buffer. The lysates, supernatant, flow-through and Ni beads were analyzed by SDS-PAGE (see figure 4.2).

In contrast to previous observations, expression at 37°C did not lead to major protein

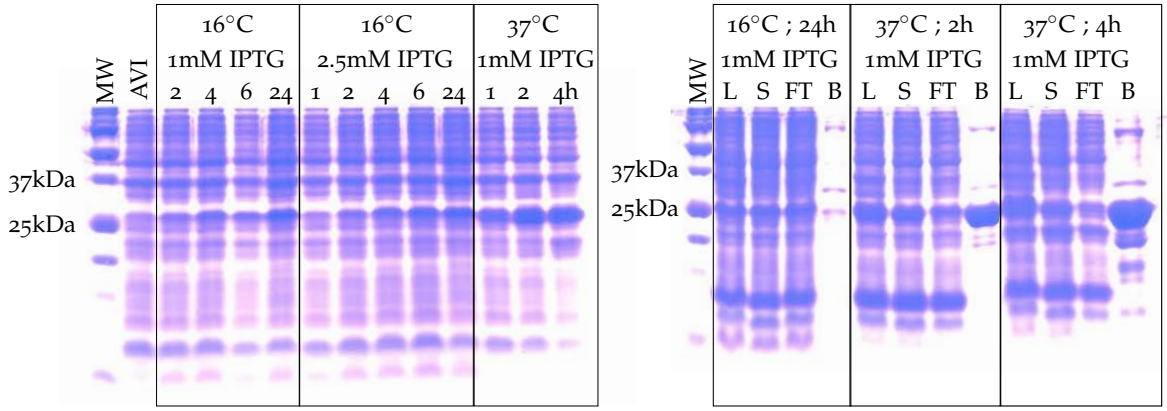


Figure 4.2 – *SDS-12.5% polyacrylamide gels showing CoblAB overexpression. On the left : bacterial extract. On the right : L=lysate; S=supernatant; FT=flow-through; B=Ni beads after incubation with the supernatant*

aggregation. Binding to the Ni beads was better for short expression times at 37°C than for longer expression times at 16°C. Truncated forms of the protein were only observed after 4h of induction at 37°C.

CoblAB expression was thereafter induced by 1mM IPTG at 37°C for 2 hours in LB medium. Similar expression tests in M9 medium showed that optimal expression conditions in this medium were 2mM IPTG at 37°C for 6 hours.

4.1.4 Protein purification

CH1 and CH2 were purified as previously described (Didry et al. 2012).

Briefly, cells were lysed using a french-press. The protein was purified using gluthathion-agarose beads. The GST was cleaved by Prescission protease and the protein was further purified by gel-filtration.

A different purification protocol was used for N-WASPV2 and CoblAB. See appendix for a detailed description of the composition of the buffers.

A cell pellet was thawed on ice and resuspended in lysis buffer. Cells were lysed using a cell disruptor (1.5kbar). Cell debris were removed by ultracentrifugation (100 000g, 40 minutes, 4°C).

The supernatant was loaded on a 5mL Histrap column equilibrated with buffer A. The column was washed with buffer A and 5% buffer B (50mM imidazole). The protein was eluted by an imidazole gradient from 50 to 500mM.

Fractions containing the protein were pooled and loaded on a desalting columns equilibrated with buffer C. The eluted protein was pooled and concentrated by ultrafiltration. The N-terminal tag was cleaved by incubation at 4°C, overnight with the PreScission or TEV protease.

The cleaved tag, uncleaved protein and the PreScission or TEV proteases which both contain an histidine tag were removed by incubation with Ni agarose beads. Contaminants

of high affinity for nickel ions which were eluted with the protein were also eliminated at this step.

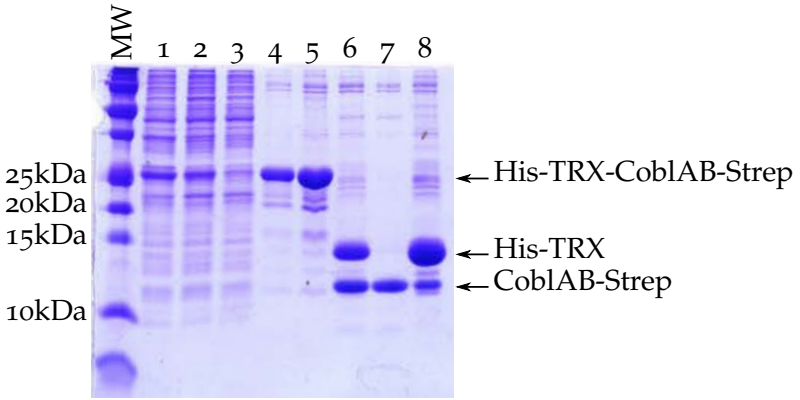


Figure 4.3 – *SDS-15% polyacrylamide gels - analysis of CoblAB purification. Lane 1 is the cell lysate. Lane 2 is the supernatant after ultracentrifugation. Lane 3 is the flow-through of Histrap column. Lanes 4 and 5 are the pooled fractions after Histrap purification and desalting column respectively. Lane 6 shows PreScission cleavage efficiency. Lane 7 is the supernatant after incubation with Ni beads. Lane 8 shows the proteins bound to Ni beads.*

To increase the protein life-time in the spectrometer, the protein was further purified by gel filtration on a preparative Superdex75 column equilibrated with G-buffer. The fractions containing the protein were pooled, concentrated, frozen in liquid nitrogen and stored at -20°C.

For CoblAB, final sample concentrations were estimated by measuring absorption at $\lambda = 280nm$ ($\epsilon_{280} = 6990 L mol^{-1} cm^{-1}$). N-WASPV2 and the chimera do not have any aromatic residues and do not absorb at $\lambda = 280nm$. For these proteins, the concentrations were estimated by a BCA assay using a BSA solution as reference.

The C-terminal Streptag of CoblAB was not used for affinity purification because an additional purification step on a Streptrap column led to a dramatic decrease in the purification yield. SDS-PAGE analysis and NMR spectra confirmed that the protein was not truncated at its C-terminus.

Because of free cysteines in position 34 and 66 respectively, N-WASPV2 and CoblAB tend to dimerize. A reducting agent (DTT or TCEP) was added before NMR experiments.

4.2 Actin

As mentioned in the introduction, actin cannot be expressed in bacterial systems. Actin used during this study was either purified from rabbit muscles or expressed in insect cells using the baculovirus/Sf9 system.

4.2.1 Tissue-purified rabbit actin

Skeletal α-actin was prepared by Bérengère Guichard (LEBS) by purification from rabbit muscle acetone powder and isolation in the CaATP-G-actin form by chromatography through Sephadex G-200 in G buffer (Spudich and Watt 1971, Macleanfletcher and Pollard 1980).

4.2.2 Recombinant drosophila actin

4.2.2.1 Proteins

Recombinant actin can be overexpressed in insect cells using the baculovirus/Sf9 system, and purified with high yield (Joel et al. 2004, Rould et al. 2006). Two actin constructs were prepared : wild-type drosophila 5C-actin (UNIPROT : P10987) and a double mutant AP-actin (A204E/P243K). These two mutations render actin nonpolymerizable (Joel et al. 2004). This particular actin mutant was previously used for structural studies (Rould et al. 2006) and in particular to investigate the structural organization complexes formed between WH2 repeats of Spire and actin (Ducka et al. 2010, Sitar et al. 2011). We planned to use the AP-actin mutant to study interaction of actin with nucleating WH2 repeats of Cobl without risking actin polymerization in the NMR tube.

Synthetic genes coding for 5C-actin and AP-actin were ordered from GeneCust. Codons were optimized for expression in insect cells. An N-terminal double Streptag was added for affinity purification followed by a PreScission cleavage site for tag removal.

Actin genes were inserted in a pVL1393 vector under a polyhedrin promoter using the BamHI and EcoRI restriction sites.

4.2.2.2 Virus preparation

Protein expression using the baculovirus Sf9 system requires the preparation of a highly concentrated viral stock.

The Baculovirus genome is too large to directly insert foreign genes easily. So the recombinant baculovirus was generated by co-transfection in Sf9 cells of the baculovirus transfer vector pVL1393 containing the actin gene and of a linearized BaculoGold DNA. The linear baculovirus DNA contains a lethal deletion and does not code for a viable virus. Homologous recombination inside Sf9 cells between the linear DNA and the transfer vector rescues the lethal deletion and produces a circular viable baculovirus DNA. Seven days after transfection, cells were pelleted and a solution containing only viable recombinant viruses was collected.

Unique virus clones were selected by an end-point dilution assay. Several clones were amplified once : Sf9 cells were infected at a low multiplicity of infection of number of viral particle per insect cell (MOI<1), incubated at 28°C for seven days, cells were pelleted and the supernatant was collected. Protein expression was evaluated by western blot and the best clone was amplified twice. Viral titer was evaluated by end-point dilution assay and

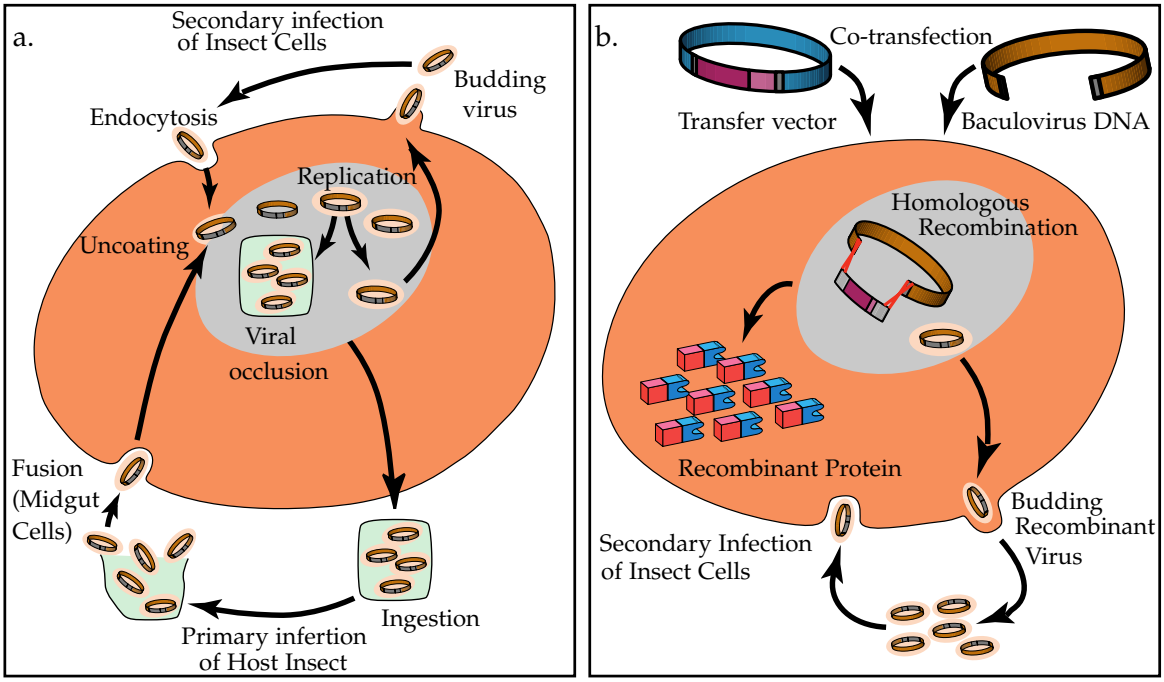


Figure 4.4 – *In vitro and in vivo life cycle of the baculovirus (from the Baculovirus Expression Vector System Manual). a. In vivo viruses exist as budded virions and occluded viruses. Occluded viruses are not necessary for secondary infection of the same insect. b. The polyhedrin or p10 protein genes are expressed during the late phase of the baculovirus cycle to form occluded viruses. In vitro, one of these genes is replaced by our gene of interest. The recombinant baculovirus is repared by co-transfection of a linear baculovirus DNA containing a lethal deletion and a transfer vector containing the gene of interest in insect cells. Homologous recombination forms a viable recombinant baculovirus containing the gene of interest which will replicate in the cell and form budding viruses. At the late phase of this recombinant virus cycle, the protein of interest is expressed instead of polyhedrin of p10 protein.*

should be above 1.10^8pfu/mL. The final viral stock was stored at 4°C and was stable for at least six months.

4.2.2.3 Protein expression

Sf9 cells were grown in Insect-Express medium (Lonza) supplemented with fetal calf serum (FCS). Sf9 cells were gently pelleted, the supernatant was discarded and the pellet was resuspended in the appropriate amount of viral stock to obtain a MOI of five. The suspension was gently shaken during an hour at room temperature by an horizontal rotative shaker. The cells were transferred into a spinner flask and culture medium supplemented with FCS and gentamicine was added to reach a cell density of 3.10^6cells/mL The culture was incubated at 28°C with shaking for three days. The cells were pelleted by centrifugation, the pellet was rinsed with PBS buffer, cells were pelleted and the pellet was collected for protein purification.

4.2.2.4 Protein purification

See appendix for a detailed description of the composition of the buffers.

A cell pellet was thawed on ice and resuspended in lysis buffer. Cells were lysed by incubation in lysis buffer on ice for 45 minutes. Cell debris were removed by centrifugation (30 000g, 40 minutes, 4°C).

The supernatant was dialysed versus buffer A overnight. The solution was clarified and loaded on a 1mL Streptrap column. The column was washed with buffer A and the protein was eluted with buffer B containing 2.5mM of desthiobiotin.

Fractions containing the protein were pooled, concentrated by ultrafiltration and further purified on a preparative Superdex75 column equilibrated with G buffer.

Fractions containing pure actin were pooled, concentrated by ultrafiltration (10 000kDa cutoff) and stored on ice at a final concentration below 50μM.

The tag could not be cleaved by the PreScission protease without precipitating the protein. Therefore recombinant actins used in our experiment have a 42-residue-long extension at the N-terminus composed of the double Streptag and PreScission cleavage site.

4.2.3 Isotopic labeling of actin

ESF921 medium depleted in all amino acids (ESF921-Δaa) was purchased from Expression Systems. The dried ^{15}N-labeled yeast autolysate was developed and provided by BioSpringer. 10 grams of ^{15}N-labeled yeast autolysate and 0.27 gram (5mM) of ^{15}N-labeled ammonium chloride were dissolved in 1 liter of ESF921-Δaa medium. The medium was sterilized by filtration on 0.22μm filters.

Protein expression was carried out as described in the previous paragraph. Sf9 cells were grown in Insect-Xpress medium and transferred in the labeled medium after infection by the baculovirus. No adaptation to the YA-ESF-Δaa medium was required.

4.2.4 Formation of the actin/WH2 complexes

NMR characterization of actin/WH2 complexes requires to work at protein concentrations above the critical concentration for actin polymerization. Actin filaments formation leads to increase in sample viscosity and formation of a gel and prevents the formation of complexes between actin monomers and WH2 repeats. It is therefore essential to avoid actin polymerization throughout sample preparation.

The critical concentration for Ca-ATP-actin polymerization at 4°C is about 70-100μM. A 50μM G-actin solution is mixed with a concentrated solution of CH1/2, NWASPV2 or CoblAB. WH2 repeats are added in excess to avoid presence of free actin monomers in solution, at least 1.1 chimera equivalents or 0.6 NWASPV2 or CoblAB equivalents. The solution is incubated on ice to allow complex formation and further concentrated by ultrafiltration (3 000kDa cutoff). The solution is homogenized every 10 minutes to avoid

filament formation. Final concentrations were 150-300μM for WH2 repeats are therefore about 150-300μM and 300-600μM for actin bound to single domains and bidomains respectively.

Biochemical characterization 5

5.1 Thermal Shift Assay

Thermal shift assay (TSA) or differential scanning fluorimetry is a method for evaluating the thermal stability of a protein (Boivin et al. 2013). For this assay, the target protein solution is mixed with an environment-dependent hydrophobic fluorescent dye, SYPRO Orange, which interacts with hydrophobic regions of proteins in a non-specific manner. If the protein is folded and stable, most of the hydrophobic regions are buried and not accessible to the dye which remains free in solution and fluorescence is quenched by interaction with water. As the temperature increases and the protein unfolds, hydrophobic regions become accessible to the fluorescent dye and the fluorescence signal increases as shown of figure 5.1. After complete denaturation, fluorescence decreases as the temperature increases because of protein aggregation and dye dissociation or decrease of the dye quantum yield with temperature.

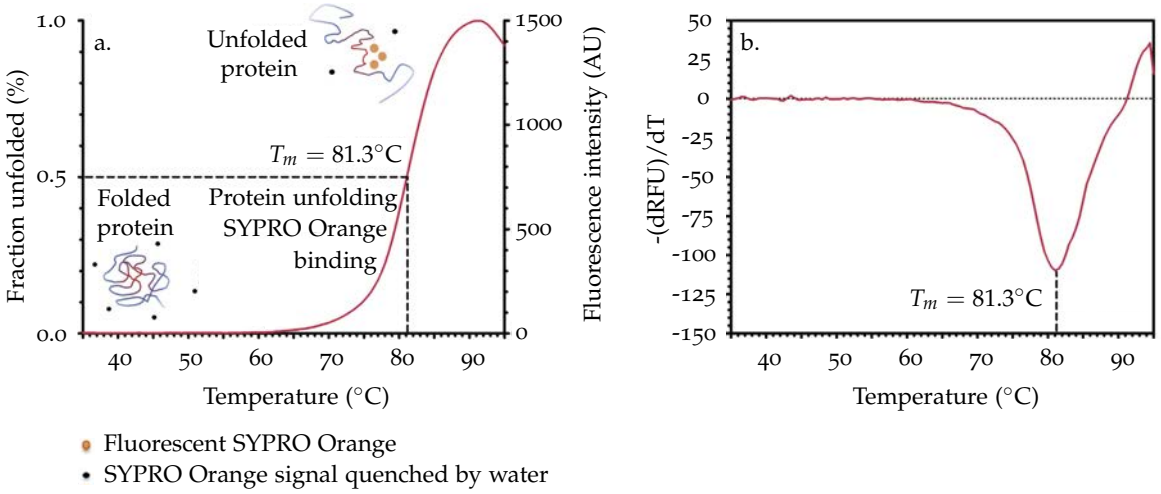


Figure 5.1 – *Typical thermal shift assay adapted from Boivin et al. (2013). a. Raw fluorescence signal showing unfolding of glucose isomerase in 100 mM Tris–HCl pH 8.5. The melting temperature correspond to the inflection point of the fluorescence curve (a.) or minimum of the first derirative curve (b.) where 50% of the protein is unfolded.*

This assay can be used to screen buffer conditions (pH, salt...) in order to increase

protein stability for structural studies (Ericsson et al. 2006, Boivin et al. 2013) or to quantify interaction with ligands, proteins or small molecules (Lo et al. 2004, Niesen et al. 2007).

Thermal shift assays were performed by Eric Jacquet and Naima Nhiri for recombinant 5C and AP-actins in various conditions. Additional controls in buffer solutions in absence of protein were necessary to make sure no buffer component interfered with the fluorescence of SYPRO Orange.

5.2 Protein-protein interactions

Isothermal titration calorimetry (ITC) is a widely used method to quantitatively analyze protein-protein interactions and does not require protein labeling. However, this method has low sensitivity, requires high amounts of material and concentrations at which actin polymerizes. Surface plasmon is another method to study protein protein interaction which requires immobilization of one of the partners on a sensor chip. We analyzed actin/WH2 repeats interactions using fluorescence based methods which are highly sensitive and because actin can easily be labeled with a fluorophore.

5.2.1 Actin labeling with a fluorophore

Fluorescence based methods require covalent coupling of the protein to a fluorophore, classically using the particular reactivity of lysine primary amine side chains or cysteine thiol side chains.

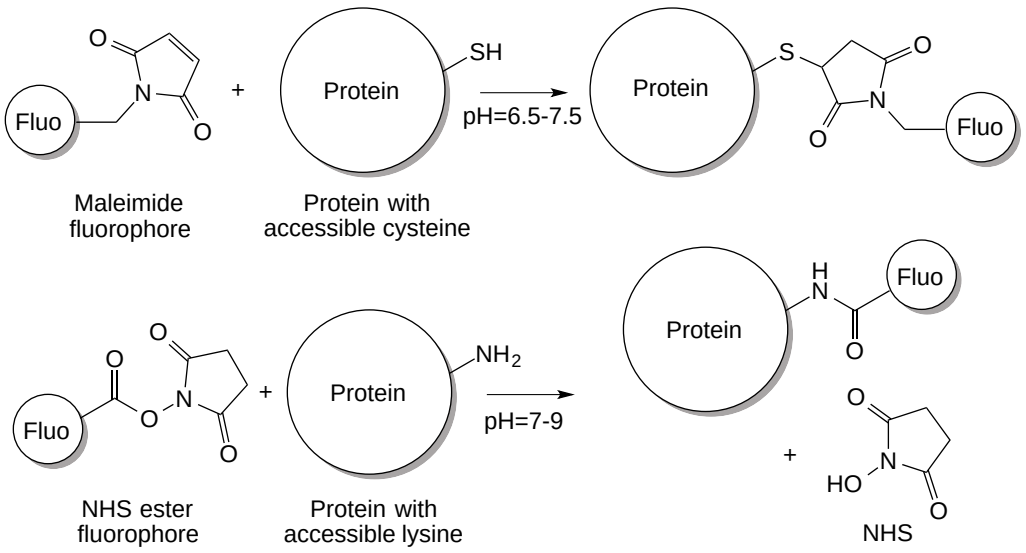


Figure 5.2 – *Maleimide-cysteine coupling & NHS-lysine coupling*

Preliminary tests on lysine labeled actin showed little fluorescence or MST signal change upon addition of CH2, therefore further experiments were carried out with cysteine labeled actin.

Rabbit skeletal α-actin has a reactive cysteine in position 374, near the C-terminus (in green on figure 5.3) that can be labeled with various fluorophores. Labeling reaction is usually carried out on F-actin filament, labeled actin filaments are then pelleted and depolymerized to obtain the labeled monomers. This method allows to make sure that labeled actin retains its polymerizing properties. However, as we wanted to compare binding of CH2 with polymerizable actin and with the non polymerizable actin mutant, all coupling reactions were carried out on actin monomers solutions in order to have comparable labeling reaction efficiency for the different constructs.

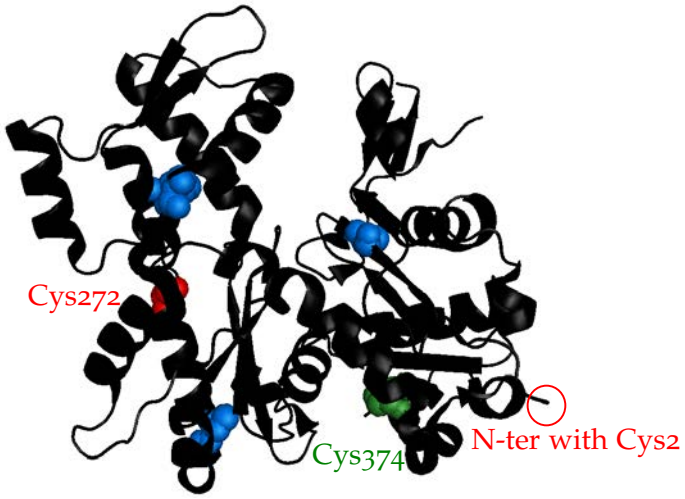


Figure 5.3 – *Crystal structure of drosophila Ca-ATP AP-actin (PDB:2HF4) with highlighted cysteine residues. In green : reactive cysteine 374. In blue : other cysteines of rabbit and drosophila actins. In red : additional cysteines of drosophila recombinant actins.*

Skeletal α-actin and recombinant 5C and AP-actins were labeled with the Monolith NT.115 protein labeling kit (NanoTemper Technologies, Germany) using the maleimide (cysteine-reactive) NT-495 blue fluorescent dye. The saline reaction buffer of the kit was replaced by a HEPES buffer, pH=7, 0.1mM $CaCl_2$, 0.2mM ATP to avoid actin polymerization and the reaction was carried out overnight in the dark at room temperature.

Labeling efficiency, evaluated by comparison of the absorbance of the protein at 280nm and of the absorbance of the fluorophore at 495nm, was 50% for rabbit actin and 70% for 5C-actin and AP-actin. Recombinant drosophila 5C and AP-actins have an accessible cysteine in position 272, not present in rabbit skeletal α-actin and an unprocessed N-terminus (because of the tag) with a cysteine in position 2 (in red on figure 5.3). These additional cysteines might also be reactive explaining the apparent "higher" labeling efficiency.

5.2.2 Microscale thermophoresis (MST) experiments

5.2.2.1 General principle

Thermophoresis is the movement of mobile particles in a temperature gradient which creates a concentration gradient (Seidel et al. 2013). At steady state, for a given spatial

temperature difference, the molecular flows are due to counterbalanced mass diffusion and thermophoretic flow :

$$j = -cD_T grad(T) - Dgrad(c) = 0 \tag{5.1}$$

where c in the molecule concentration, T is the temperature, D is the diffusion coefficient and D_T is the thermal diffusion coefficient.

The concentration difference between the 'hot' and 'cold' zones of the temperature gradients is then given by the following equation :

$$\frac{c_{hot}}{c_{cold}} = exp(-S_T \Delta T) = exp(-\frac{D}{D_T}\Delta T) \tag{5.2}$$

where S_T is the Soret coefficient.

This concentration difference depends on the size, charge, conformation of the molecule and on the hydration shell. At least one of these parameters is susceptible to be affected upon interaction therefore MST enables to probe a wide range of interactions. On the other hand, opposite effects on different parameters can complicate data interpretation.

In MST experiments, an infrared laser creates an important temperature gradient at microscopic scale ($\Delta T = 2 - 6K$). The microscopic scale is essential for fast experiments because thermophoresis is a diffusion limited process. Therefore, this method requires very low sample volumes of 20μL per analysis.

5.2.2.2 Binding assays and data analysis

MST experiments to investigate protein-protein interactions were carried out as follows. One protein was labeled with a fluorophore. Final concentration of the labeled protein should be of the order of magnitude of the expected K_D or lower. Serial dilutions of the partner were prepared, added to the labeled protein and the solutions were loaded into capillary tubes. Fluorescence was measured for the initial state, during IR laser irradiation and after the laser was turned off as shown on figure 5.4. The LED power for excitation of the fluorophore was optimized to get an initial fluorescence between 200 and 2000. All experiments were carried out with MST powers of 20 and 40% for IR irradiation.

Binding assays were performed in G-buffer (5mM tris, pH=7.8, 0.1mM $CaCl_2$, 0.2mM ATP) with 0.05% Tween-20 on a Monolith NT.115 Microscale Thermophoresis device using standard treated capillaries (NanoTemper Technologies). The final actin concentration was 50nM which is below critical polymerization concentration both in G-buffer (no salt) and F-buffer (150mM KCl). Here, experiments were carried out in G-buffer because actin/WH2 binding affinities are in general lower in F-buffer so that much higher WH2 repeat concentrations are required to build the titration curves with risks of WH2 repeats aggregation. Actin was titrated by 1:1 series dilutions of CH2 at final concentrations ranging from 0.15nM to 40μM. Tween-20 was added to the buffer as recommended by the manufacturer to avoid aggregation or sticking to the capillaries.

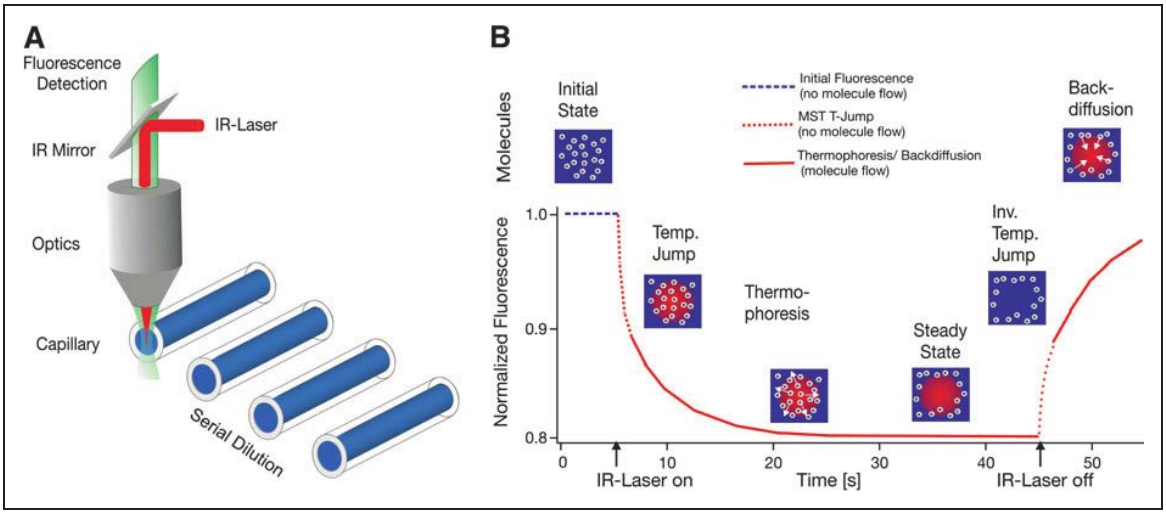


Figure 5.4 – *Typical MST experiment (Jerabek-Willemsen et al. 2011). A : experimental setup. B : MST signal for a given capillary. The solution is homogeneous at initial state. IR irradiation and sample heating lead to thermophoretic motions out of the heated zone and decrease in the fluorescence signal until a plateau is reached (steady-state). Turning off the laser leads to fast sample cooling and molecules back diffusion.*

All binding assays were repeated 3 times to evaluate experimental error.

The thermophoresis signal taking into account the temperature jump is the fluorescence after thermodiffusion divided by the initial fluorescence. Change in the initial fluorescence may also be observed if the fluorophore's environment is modified upon binding.

In our case, initial fluorescence was modified upon CH2 addition with fluorescence intensity increase up to 70%. The MST curves showed a more complex biphasic behavior (see subsection 7.1.2). Therefore we determined the binding affinity from the fluorescence curves using a simple two-states exchange model.

Fluorescence curves were fitted to the following equation with Matlab using non-linear fitting and Monte-Carlo runs for error estimates.

$$F([CH2]) = F_U + (F_B - F_U)\frac{([CH2] + [actin] + K_d) + \sqrt{([CH2] + [actin] + K_d)^2 - 4[CH2][actin]}}{2[CH2]} \tag{5.3}$$

F_U and F_B are the raw fluorescence intensities for free actin and actin bound to CH2 respectively. K_d is the dissociation constant of the actin:CH2 complex.

NMR spectroscopy

6

All NMR experiments were realized on Bruker spectrometers operating at ^{1}H frequencies of 600, 700, 800 and 950MHz. All spectrometers, expect for the 700MHz, were equipped with a TCI cryoprobe with z-gradients.

6.1 Protein backbone assignment

The first step in any protein NMR study is the assignment of the protein backbone which enables one to associate resonances of the spectra to the nucleus of a particular atom of the protein.

6.1.1 General strategy

Modern protein assignment is based on multidimensional heteronuclear NMR experiments on uniformly ^{15}N-^{13}C-labeled sample. Strong J couplings between ^{13}C, ^{15}N and their attached protons (see figure 6.1) enables efficient magnetization transfer to establish the connectivity between NMR active nuclei.

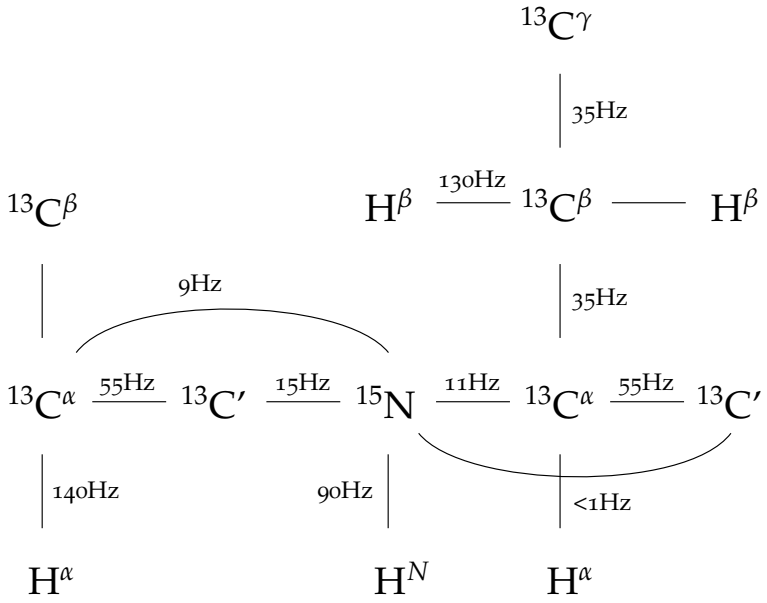


Figure 6.1 – *Average values of ^{1}J and ^{2}J scalar coupling constants within the backbone of ^{15}N-^{13}C-labeled proteins (adapted from Cavanagh et al. (2007))*

Most amino acids have a unique amide HN group, except for proline wich has none and asparagine and glutamine which also have an amide side chain. In classic 3D experiments for protein assignment, HN groups are correlated with another type of nucleus on a sequential or bidirectional manner. These experiments are named by the nuclei involved in the transfer, with the chemical shifts of the nuclei in parenthesis not being mapped. For example, on the one hand, the HNCO experiment shows correlation between $^{1}H^{N}$, ^{15}N resonances of a residue i and the $^{13}C'$ resonance of the previous residue i-1. On the other hand, the HN(CA)CO experiment correlates $^{1}H^{N}$ and ^{15}N resonances of a residue i with the $^{13}C'$ resonances of the previous residue i-1 but also of the same residue i via the non-edited aliphatic carbons $^{13}C^{\alpha}$.

^{1}H-^{13}C strips extracted from the 3D spectra at a given 15N-frequency are aligned to allow sequential assignment as shown in figure 6.2. Some amino acid types (alanines, glycines, serines and threonines) can be easily identified thanks to their remarkable $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ chemical shifts and are used as starting points in sequential assignment.

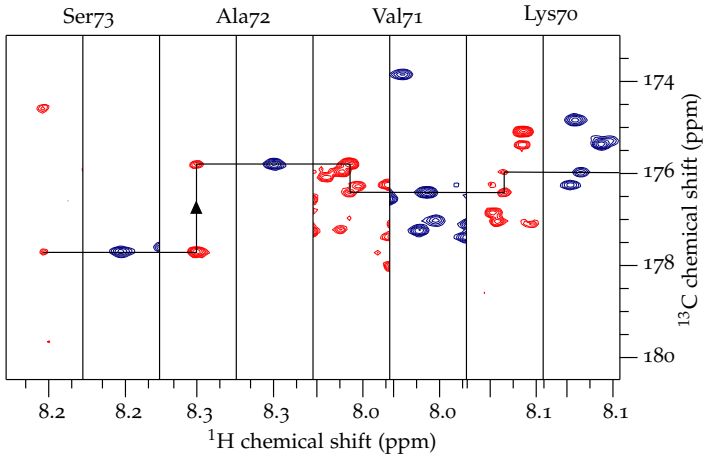


Figure 6.2 – *^{1}H ^{13}C strips from HNCO (red) and HN(CA)CO (blue) experiments showing NMR assignment of residues 70 to 73 of CoblAB. For a given residue, the H^{N} and ^{15}N resonances are correlated with the $^{13}C'$ resonances of the same and previous residues in the HN(CA)CO and with the $^{13}C'$ resonance of the previous residue in the HNCO.*

6.1.2 NMR experiments

The backbones of CH1 and CH2 were assigned by François-Xavier Cantrelle during is PhD (Cantrelle 2010).

The backbones of CoblAB and N-WASPV2 were assigned using experiments listed in table 6.1. The protein concentration was 400-500μM. The pH was ajusted to 7 and the temperature set at 278K to slow down water amide hydrogen exchange.

Multidimensional NMR experiments are usually collected on a grid of equally-spaced points in the time space of indirect dimensions. Here, NMR spectra were recorded using non-uniform sampling (NUS) and acquisition time was divided by four to five compared to traditional methods. Because of the sparse data in indirect dimensions, spectra cannot

NMR experiment	Observed nuclei
HNCO	H^N(i), ^{15}N(i), ^{13}C′(i-1)
HN(CA)CO	H^N(i), ^{15}N(i), ^{13}C′(i-1) and ^{13}C′(i)
HNCA	H^N(i), ^{15}N(i), $^{13}C^{\alpha}$(i-1) and $^{13}C^{\alpha}$(i)
CBCA(CO)NH	H^N(i), ^{15}N(i), $^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1)
CBCANH	H^N(i), ^{15}N(i), $^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1) and $^{13}C^{\alpha}$(i), $^{13}C^{\beta}$(i)

Table 6.1 – *NMR experiments for backbone assignment of CoblAB and N-WASPV2*

be obtained by discrete Fourier transform. NMR data were processed using the qMDD software : the directly detected proton dimension was processed by Fourier transform, the complete data matrix was reconstructed using compressed sensing and finally the indirect nitrogen and carbon dimensions were processed by Fourier transform (see figure 6.3).

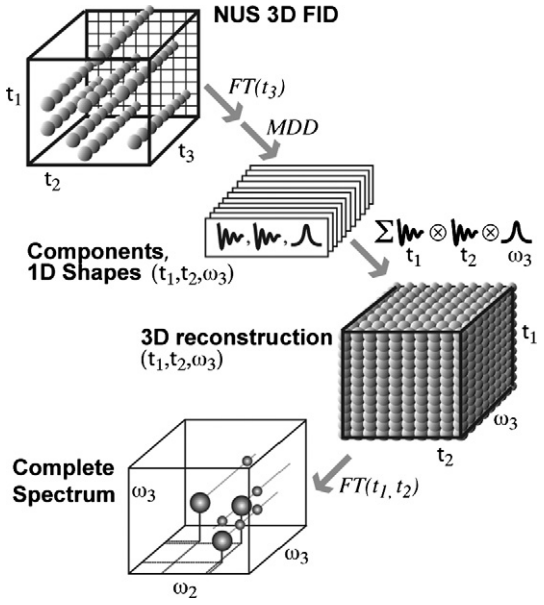


Figure 6.3 – *Processing of NMR experiments recorded with non-uniform sampling (Orekhov and Jaravine 2011)*

Spectra were analyzed and assignment was performed in CCPNMR Analysis 2.1.5 (Vranken et al. 2005).

HSQC spectra were recorded every 5K for propagation of the assignment from 278K to 298K.

6.2 Protein structure and dynamics by NMR

NMR spectroscopy is a powerful tool to achieve description of proteins structure and at atomic resolution Cavanagh et al. (2007). Many NMR observables are sensitive probes

of polypeptide structure and motions on a wide timescale ranging from picoseconds to seconds (see figure 6.4).

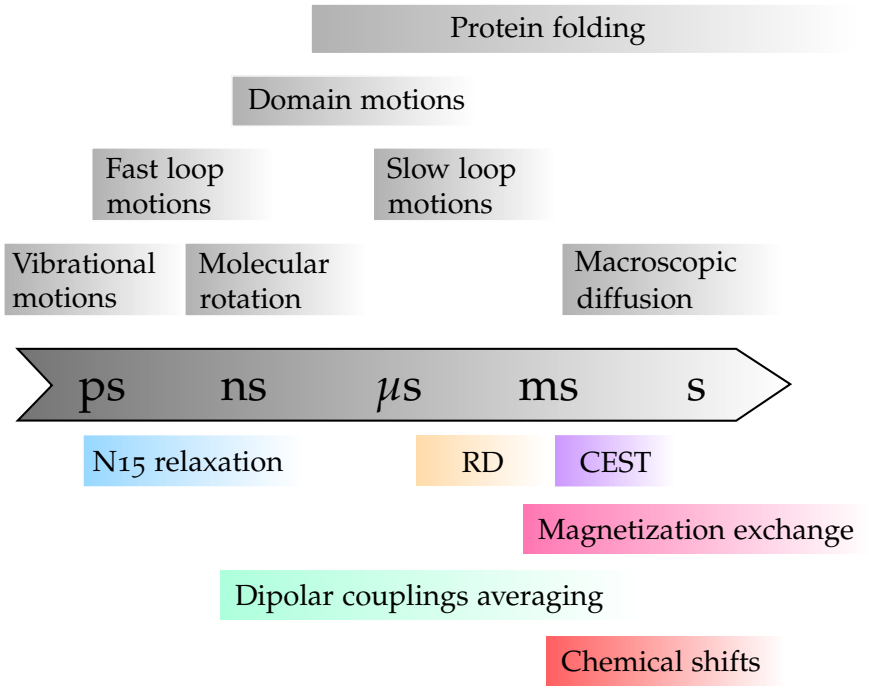


Figure 6.4 – *Time-scales of protein motions and NMR observables*

Chemical shifts and residual dipolar couplings are sensitive to the protein structure. These two NMR observables do not average over the same time range : ∼1-10ms for chemical shifts and ∼100μs for RDC. Besides, it is interesting use both parameters because some conformational degeneracies are inherent to the data type (Ozenne et al. 2012b). For example ^{13}C backbone chemical shifts enable to quantify α-helices and β-sheets but will not distinguish polyproline II and random coil conformations. This degeneracy can be raised by the use of proton and nitrogen amide chemical shifts but these are also more sensitive to experimental conditions like pH or temperature. RDC enable quantification of α-helices, extended conformers and random coil but will not distinguish β-sheet and polyproline II conformations.

Chemical exchange between amide and water protons of the solvent gives information on the accessibility of amide protons and therefore on secondary structures and their stability in the millisecond timescale.

^{15}N-relaxation rates provides information on the fast picosecond to nanosecond backbone dynamics.

Chemical exchange parameters measured by relaxation dispersion (RD) and chemical exchange saturation transfer (CEST) experiments probes conformational exchange on the micro to millisecond time scale. Relaxation dispersion is an optimal method to probe conformational exchange on a time scale ranging from 500μs to 5ms. CEST experiments enable one to characterize slower motions on the 5 to 50ms time scale.

IDPs cannot be described by a single structure but more by ensembles of interconverting flexible conformers. It should be noted that disordered does not necessarily means random coil. IDPs may form transient partial secondary structures or tertiary contacts important for their function and these constraints should be taken into account in ensemble description, for example by definition of several sub-ensemble including major secondary structure elements and determination of their respective populations (Jensen et al. 2008). Special care should be taken to define the size and composition of conformational ensembles : ensembles size should allow broad enough conformation sampling and and avoid under-definition whilst avoiding overfitting by accumulation of sub-ensembles.

NMR is a technique of choice for ensemble description of IDPs because various NMR experiments give information on residual structural elements and protein dynamics on multiple timescales at atomic resolution (Jensen et al. 2014).

6.3 Chemical shift analysis

Chemical shifts highly depend on the physico-chemical environment of the nuclei determined by structure and dynamics of the polypeptide chain and experimental conditions. α, β carbon, carbonyle, amide and α-proton and amide nitrogen backbone chemical shifts are especially sensitive to local structure. Amide hydrogen and nitrogen chemical shifts are also sensitive to experiments conditions like temperature and pH, these effects are more complicated to take into account. In the case of folded proteins, multiple methods and softwares are available to identify secondary structure elements and/or evaluate backbone dihedral angles among which CSI (Wishart and Sykes 1994), TALOS (Shen et al. 2009) and DANGLE (Cheung et al. 2010). A simple but widespread method is to compare experimental backbone chemical shifts with those expected for a random coil :

$$\Delta\delta X_{i\ obs} = \delta X_{i\ exp} - \delta X_{i\ RC} \tag{6.1}$$

The obtained $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shifts are expected to be positive in α-helices and negative in β-sheets. $^{13}C^{\beta}$ have opposite behavior which enables adjustment of the carbon chemical shifts referencing (Marsh et al. 2006). In the chemical shift index (CSI) approach, secondary chemical shifts are converted into scores of 1, -1 and 0 for α-helix, β-sheet and non structured respectively. A minimum of 3 consecutive -1 scores and 4 consecutive 1 scores are necessary to define β-sheets and α-helices respectively.

In the case of disordered proteins, the matter is not so much the identification of fully formed secondary structure elements rather than the quantification of various transient structural propensities along the sequence. Chemical shifts provide averaged structural information about conformations sampled by the protein on the ms time scale (Kragelj et al. 2013).

Because we analyze the deviation from reference values, the definition of random coil chemical shifts is crucial. This is especially important in the case of IDPs for which

deviations from random coil values are expected to be small. Random coil chemical shifts are expected values for a polypeptide chain devoid of any secondary structure. Proteins are heteropolymers so these reference values should take into account local amino acid composition are neighbors effects. Different methods reviewed by (Kragelj et al. 2013) can be classified in two general approaches : measurement of reference chemical shifts on small peptides in non-denaturing (Kjaergaard and Poulsen 2011), mildly denaturing (Wishart et al. 1995, Kjaergaard et al. 2011) and denaturing (Schwarzinger et al. 2000; 2001, Prestegard et al. 2013) conditions or derivation of random coil chemical shift from a library of chemical shifts in loop regions of folded proteins (Wang and Jardetzky 2002a;b, Zhang et al. 2003, De Simone et al. 2009) or of chemical shifts in known non-structured IDPs (Tamiola et al. 2010). Another possibility is to denature the protein under study to extract random coil chemical shifts intrinsic to this protein (Modig et al. 2007). Random coil chemical shifts determined in denaturing conditions do not take into account effects of extreme pH and influence of denaturants like hydrogen bonds between urea and the protein backbone (Lim et al. 2009) and are less likely to agree with experimental value for natively random coil proteins. Programs predicting transient secondary structure in IDPs rather use random coil chemical shifts determined in native experimental conditions.

In this study two programs were used for chemical shifts analysis : SSP (Marsh et al. 2006) and δ2d (Camilloni et al. 2012).

The SSP program uses $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts and returns a score ranging from -1 to 1 for fully formed β-sheet and α-helix respectively. Intermediate score quantify the fraction of conformers in these secondary structures.

The SSP score for a given residue i is calculated using the following equation and takes into account effects of the two preceding and following residues :

$$SSPi = \frac{\sum_{j=i-2}^{i+2} \sum_{X} \begin{cases} \frac{|\Delta\delta X_{j\ obs}|}{\sigma_{j\alpha}} & if(\Delta\delta X_{j\ obs}) \times (\Delta\delta X_{j\ \alpha}) > 0 \\ -\frac{|\Delta\delta X_{j\ obs}|}{\sigma_{j\beta}} & if(\Delta\delta X_{j\ obs}) \times (\Delta\delta X_{j\ \beta}) > 0 \end{cases}}{\sum_{j=i-2}^{i+2} \sum_{X} \begin{cases} \frac{|\Delta\delta X_{j\alpha}|}{\sigma_{j\alpha}} & if(\Delta\delta X_{j\ obs}) \times (\Delta\delta X_{j\ \alpha}) > 0 \\ \frac{|\Delta\delta X_{j\beta}|}{\sigma_{j\beta}} & if(\Delta\delta X_{j\ obs}) \times (\Delta\delta X_{j\ \beta}) > 0 \end{cases}} \tag{6.2}$$

$\Delta\delta X_{j\ obs}$ is the secondary chemical shift of the nucleus X of residue j. $\Delta\delta X_{j\alpha/\beta}$ are the average secondary chemical shifts of the nucleus X ($^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ of $^{13}C'$) of residue j fully α-helical or β-sheet and $\sigma_{j\alpha/\beta}$ are the standard deviations of $\Delta\delta X_{j\alpha/\beta}$.

The average random coil, α-helix and β-sheet chemical shift are taken from the Re-referenced Protein Chemical shift Database (RefDB) (Zhang et al. 2003).

The SSP method has a few drawbacks. It does not distinguish between β-sheets and polyproline II which are both extended structures with average (ϕ, psi) torsion angles of (-75°, 150°) and (-135°, 135°) respectively. Besides, if regions of the protein have both helical and extended propensities, the SSP score will only reflect the prevailing trend.

The δ2d algorithm predicts the fraction of α-helices, β-sheets, polyproline II and random coil per residue using the $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts.

Standard chemical shift values derived from a database of 1772 proteins for which both three dimensional structure and chemical shifts are known. Residues in loops, α-helices longer than three residues, β-sheets and polyproline II longer than three residues were extracted and four different chemical shifts predictors were set up for each type of nucleus which include correction factors to take into account the nature of the preceding and following residues i-1 and i+1 for prediction of random coil chemical shifts(De Simone et al. 2009), the nature of residues i-1 to i+4 for α-helices and the nature of residues i-2 to i+2 for β-sheets and polyproline II (Camilloni et al. 2012).

The following equations are used to predict the chemical shift of a nucleus of residue i of type A (Ala, Gly, etc) in random coil :

$$\delta_{iA}^{RC} = \delta_{iA}^{0} + \alpha_i^{-}\delta_{iBA}^{1} + \alpha_i^{+}\delta_{iAC}^{1} \tag{6.3}$$

in a α-helix

$$\delta_{iA}^{\alpha} = \delta_{iA}^{0} + \alpha_i^{-}\delta_{iBA}^{1} + \alpha_i^{+}\delta_{iAC}^{1} + \alpha_i^{2+}\delta_{iAxC}^{1} + \alpha_i^{3+}\delta_{iAxxC}^{1} + \alpha_i^{4+}\delta_{iAxxxC}^{1} \tag{6.4}$$

in a β-sheet

$$\delta_{iA}^{\beta} = \delta_{iA}^{0} + \alpha_i^{2-}\delta_{iBxA}^{1} + \alpha_i^{-}\delta_{iBA}^{1} + \alpha_i^{+}\delta_{iAC}^{1} + \alpha_i^{2+}\delta_{iAxC}^{1} \tag{6.5}$$

and in a polyproline II :

$$\delta_{iA}^{PPII} = \delta_{iA}^{0} + \alpha_i^{2-}\delta_{iBxA}^{1} + \alpha_i^{-}\delta_{iBA}^{1} + \alpha_i^{+}\delta_{iAC}^{1} + \alpha_i^{2+}\delta_{iAxC}^{1} \tag{6.6}$$

where B and C are amino acid nature of preceding and following residues. Correction factors depend on the residue types and relative positions of residues A and B or C.

These predicted chemical shifts are then combined to extrapolate the fraction of the four different types of conformation for every residues by comparison with experimental chemical shifts.

6.4 Residual dipolar couplings

Dipolar interactions between two NMR active nuclei are governed by the following equation :

$$D_{ij} = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2 r_{ij}^3}\left\langle\frac{3cos^2\theta - 1}{2}\right\rangle \tag{6.7}$$

where r_{ij} is the distance between the two nuclei and θ is the angle between the vector of the two nuclei and the static magnetic field.

In solution NMR experiments, dipolar couplings are averaged to zero, because of the random orientation and tumbling of the molecule. But in some aligning media such as

dilute liquid crystals, all orientations of the protein and θ angles are no longer equiprobable and residual dipolar couplings can be measured. These RDCs can be used to refine protein structures in the case of folded proteins. In the case of IDPs, they give information on residual secondary structures and long range contacts.

If the aligning medium is parallel to the static magnetic field, isolated α-helices and β-sheet structures will be preferentially oriented in this direction leading to negative RDCs for extended regions ($\theta \simeq 90°$) and positive RDCs for α-helices ($\theta \simeq 0°$) as shown on figure 6.5.

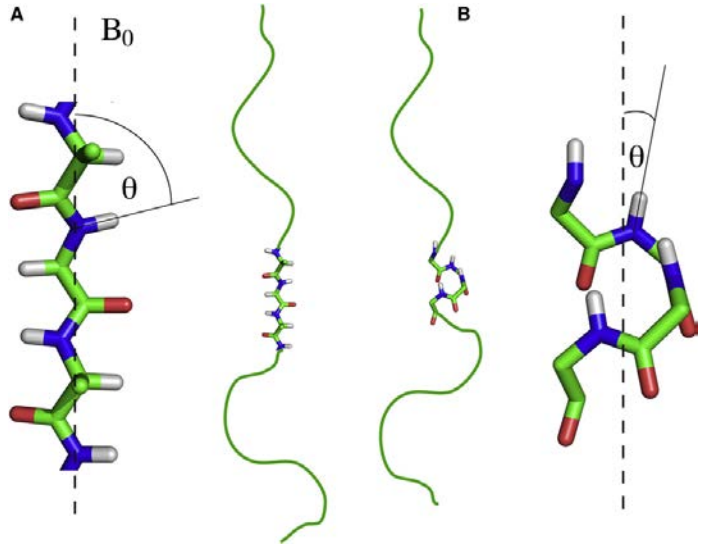


Figure 6.5 – *Preferential orientations of α-helices (A.) and extended regions (B.) of an IDP in solution with a weakly aligning medium in the direction of the static magnetic field B_0 (Jensen et al. 2009)*

6.4.1 Sample preparation

RDC measurements were carried out at 278K and pH=7 in G-buffer.

Aligning media for protein RDC measurements should be miscible with water, have a stable anisotropy over the duration of NMR experiments (a few days), not interact with the protein of interest and induce only moderate anisotropy to avoid major line broadening. Commonly used media include bicelles(Tjandra and Bax 1997), solution of filamentous phages Pf1(Clore et al. 1998, Hansen et al. 1998), lamellar phases composed of n-alkyl-poly(ethylene glycol)/n-alkyl alcohol mixtures or Otting media (Ruckert and Otting 2000), strained polyacrylamide gels (Ishii et al. 2001, Chou et al. 2001). In aqueous solutions, phospholipid mixtures form discotic micelles of about 40Å of thickness and a few hundreds of angstrom of diameter that align in the magnetic field with their normal orthogonal to the magnetic field. However, these bicelles are only stable on a narrow pH and temperature range and are sensitive to hydrolysis. Filamentous phages Pf1 are rods of about 20 000Å of length and 60Å of diameter and form a liquid crystalline phase parallel to the static magnetic field. The degree of alignment is tunable by adjustment of phage

concentration. The coat protein on the surface of the phage is strongly negatively charged (pI≃4.0) therefore this method may not be suitable for alignment for polyelectrolyte IDPs such as WH2 repeats.

Strained polyacrylamide gels are probably the most chemically inert aligning medium and are less likely to interact with proteins in solution. However, sample preparation is more constraining, alignment tensor are less homogeneous and less easily tuned and the protein sample is harder to recover from hydrated gels.

n-alkyl-poly(ethylene glycol)/n-alkyl alcohol (PEG/alcool) mixures form lamellar (L_α) phases that orient with their surface parallel to the magnetic field. These mixtures are stable in a wide pH and salt range and different kind of PEG/alcohol form lamellar phases for temperatures ranging from -5°C to 40°C as shown on figure 6.6. Sample preparation is fast, easy reproducible, the degree of alignment is tunable by adjustment of added (PEG/alcool) quantity and the protein sample is recovered by purification on a desalting column.

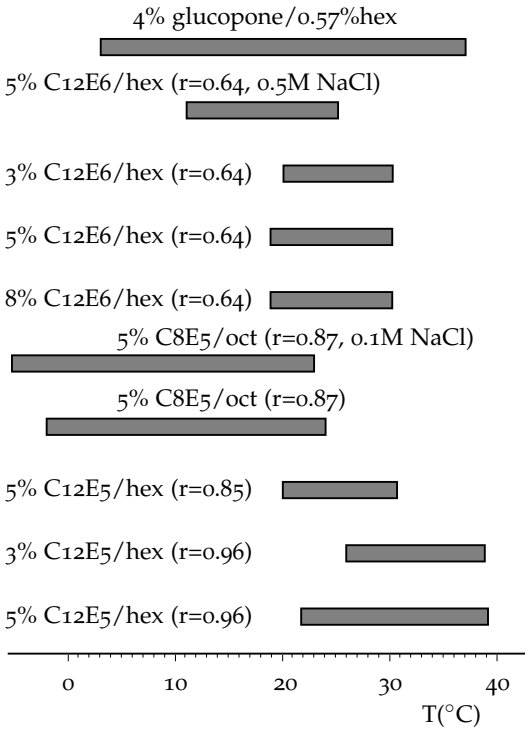


Figure 6.6

CH1, CH2 and N-WASPV2 were aligned in a dilute liquid crystalline phase of pentaethylene glycol mono-*n*-octyl ether (C8E5) and 1-octanol, more stable at low temperature (278K) than the most commonly used pentaethylene glycol mono-*n*-dodecyl ether (C12E5) / 1-hexanol liquid crystalline phase (see figure 6.6. The NMR sample was mixed with $5\%_{vol}$ of C8E5 and 1-octanol was gradually added until the phase transition was observed (about $2\%_{vol}$ of 1-octanol were required).

CoblAB showed evidences of interaction with the C8E5/octanol mixture : line broad-

ening and decrease in peak intensities. Therefore for this protein, RDCs were measured in strained polyacrylamide gels prepared using the apparatus described on figure 6.7. 6% polyacrylamide gels were polymerized in the cylindrical chamber, cut in cylinders of 1cm of height, dialyzed in water to eliminate the buffer and unpolymerized acrylamide and dehydrated at room temperature. Gels were rehydrated in the cylindrical chamber by adding 300μL of protein solution containing 10% D_2O to the dry gel. The hydrated gel was inserted into an open-ended NMR tube of smaller diameter using the apparatus shown on figure 6.7.

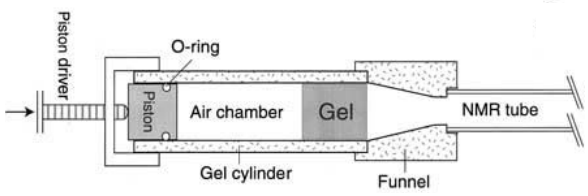


Figure 6.7 – *Apparatus for strained polyacrylamide gels preparation (Chou et al. 2001)*

Anisotropy of the aligned sample is evaluated by measurement of the HDO quadripolar splitting of water on a one-dimensional deuterium spectrum. HDO quadripolar splitting was between 20 and 40Hz for C8E5/octanol samples and between 10 and 12Hz for strained polyacrylamide gels.

6.4.2 Pulse sequences

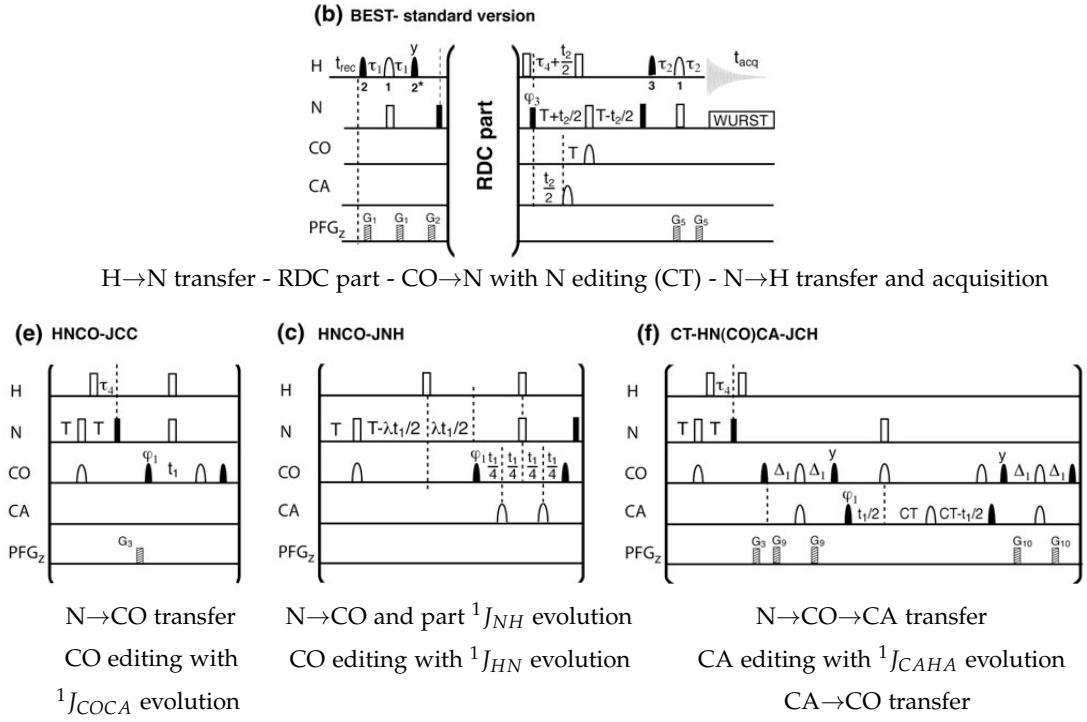
$^1D_{NH}$, $^1D_{C\alpha H\alpha}$ and $^1D_{C'C\alpha}$ RDC were measured using the four BEST-type pulse sequences (Rasia et al. 2011) detailed in table 6.2 and figure 6.8. In Band-selective Excitation Short-Transient (BEST) type experiments, amide protons are selectively excited and dipolar interactions with the unperturbed aliphatic proton accelerates longitudinal (T_1) relaxation enabling decrease of the inter-scan delay from 1s up to 0.2s (Schanda et al. 2006). Experiments were recorded in isotropic as well as in aligned media in order to remove the contribution of through-bond scalar couplings.

Pulse sequence	Measured coupling	Splitting dimension
B-HSQC-IPAP	$^1J_{NH} + ^1D_{NH}$	^{15}N
B-HNCO-JNHscale	$1.5 \times (^1J_{NH} + ^1D_{NH})$	$^{13}C'$
B-HNCO-JCOCA	$^1J_{C'C\alpha} + ^1D_{C'C\alpha}$	$^{13}C'$
B-HNCOCA-JCH	$^1J_{C\alpha H\alpha} + ^1D_{C\alpha H\alpha}$	$^{13}C^{\alpha}$

Table 6.2 – *NMR experiments for RDC measurements*

The B-HSQC-IPAP is based on a J-coupled BEST-^{1}H-^{15}N-HSQC where two in-phase and antiphase spectra are acquired on an interleaved manner. The sum and difference spectra each contain one component of the doublet to measure the ^{1}H-^{15}N coupling.

The B-HNCO-JNHscale and B-HNCO-JCOCA experiments are based on a BEST-HNCO with constant time in ^{15}N dimension and line splitting along the ^{13}C′ axis. In the



H→N transfer - RDC part - CO→N with N editing (CT) - N→H transfer and acquisition

N→CO transfer
CO editing with
$^1J_{COCA}$ evolution

N→CO and part $^1J_{NH}$ evolution
CO editing with $^1J_{HN}$ evolution

N→CO→CA transfer
CA editing with $^1J_{CAHA}$ evolution
CA→CO transfer

Figure 6.8 – *NMR BEST type pulse sequences for RDC measurements (Rasia et al. 2011)*

B-HNCO-JNHscale experiment, line splitting along the $^{13}C'$ axis is artificially increased to $1.5 \times (^1J_{NH} + ^1D_{NH})$ thanks to accordion spectroscopy : the ^1H-^{15}N coupling evolution is active during part of the $^{15}N \rightarrow ^{13}C'$ transfer and during $^{13}C'$ editing. In the B-HNCO-JCOCA experiment, the $^{13}C_\alpha$-$^{13}C'$ coupling evolution is active during $^{13}C'$ editing

The B-HNCOCA-JCH experiment is based on a BEST-HN(CO)CA with line splitting along the $^{13}C^\alpha$ axis. The $^{13}C^\alpha$-$^1H^\alpha$ coupling evolution is active during $^{13}C^\alpha$ editing which is constant time to avoid $^{13}C^\alpha$-$^{13}C^\beta$ coupling.

Residual dipolar coupling experiments were carried out at ^1H frequencies of 600MHz for CH1, CH2 and NWASPV2. Spectra were processed and analyzed using NMRPipe (Delaglio et al. 1995).

6.4.3 RDC analysis

RDCs were analyzed using the Flexible-Meccano software (Ozenne et al. 2012a) which predicts RDCs on the basis of the protein sequence and with or without imposed secondary structure propensities and tertiary contacts constraints. This algorithm generates protein ensembles for disordered proteins by randomly sampling conformers taking into account backbone dihedral angle $\{\phi/\psi\}$ preferences for each type of amino acid and avoidance of steric clashes. Backbone dihedral angle $\{\phi/\psi\}$ potential wells are derived from conformation of regions not involved in secondary structure elements from high resolution protein structures and take into account the amino acid type and the possible presence of a proline residue in position i+1.

A series of experimental parameters (RDCs, J couplings, PREs, SAXS curves) are predicted for each conformer and averaged over the ensemble.

If deviations are observed between experimental data and predicted values, additional conformational propensities can be taken into account for ensemble generation : a. local conformational propensities by backbone dihedral angle $\{\phi/\psi\}$, b. residual secondary structures (α-helices, β-sheets and polyproline II) and c. long range contacts between residues or entire regions.

6.5 Water-amide proton exchange by CLEANEX experiments

Chemical exchange rates between water and amide protons give information on the accessibility of amide accessibility to the solvent and so on the structure and dynamics of proteins.

6.5.1 NMR experiments

Proton/amide proton exchange rates were measured using the bruker CLEANEX pulse sequence (fhsqccxf3gpph) (Hwang et al. 1998).

Water is selectively excited. A CLEANEX-PM spin lock sequence is applied for a mixing time τ_m allowing magnetization transfer to amide protons by chemical exchange. The rest of the sequence is a fast HSQC with watergate water suppression.

Spectra intensity is close to zero for τ_m and increases with τ_m at rates depending on water/amide proton exchange. Several CLEANEX-HSQC are recorded with different mixing times to allow extraction of the exchange rate k_{ex}. A fast HSQC is also recorded as reference.

6.5.2 Data analysis

Proton/amide proton exchange rates are obtained by adjustment of k and R_{1A}^{app} to fit the following equation :

$$\frac{V(:tau_m)}{V_0} = \frac{k}{R_{1A}^{app} + k - R_{1B}^{app}} \times \left(exp(-R_{1B}^{app}\tau_m) - exp((-R_{1A}^{app} + k)\tau_m)\right) \tag{6.8}$$

where $V(\tau_m)$ is the peak volume for a CLEANEX mixing time τ_m. V_0 is the peak volume in the reference experiment. R_{1A}^{app} and R_{1B}^{app} are the effective longitudinal relaxations of the amide protons and water protons respectively. R_{1A}^{app} values are different from R_{1A} values determined by ^{15}N-relaxation experiments because of the magnetization pathway which includes contribute of both longitudinal and transverse relaxation.

k designates the normalized forward rate constant of the following equation :

$$NH \xrightarrow{k_{AB}=X_B \times k} H_2O$$

X_B is the fraction of protonated water in the sample.

R_{1B}^{app} is determined using a separate experiment. A serie of 1D CLEANEX-PM without water suppression are recorded and the dependence of the water signal volume on mixing times τ_m is adjusted to a monoexponetial function to extract the apparent water relaxation time.

The interscan delay does not enable complete return to equilibrium of water magnetization. Therefore, for accurate quantification of k_{AB}, one needs to determine the percentage of effectively saturated water using separate 1D experiments : equilibration is achieved by dummy scans which contain the entire CLEANEX pulse sequence, signal is then acquired after water excitation and before the CLEANEX-PM spin lock mixing time. The water signal is compared for 1D experiments with interscan delays of 1 and 30s and for mixing times of 5 and 100ms and water saturation was 80 to 90% depending on samples and temperatures. Equation 6.8 is corrected by dividing k by the fraction of saturated water.

These experimental exchange rates k_{ex} were compared with predicted values for a fully random coil protein (Bai et al. 1993, Connely et al. 1993) in the sample pH and temperature conditions. Protection factors are predicted random coil exchange rate divided by experimental exchange rates (k_{ex}^{rc}/k_{ex}^{exp}) and provide information on restriction in accessibility of amide protons.

6.6 Pico- to nanosecond dynamics by ^{15}N relaxation experiments

^{15}N relaxation experiments probe fast dynamics of proteins backbone from internal flexibility to overall tumbling on the picosecond to nanosecond timescale. These type of experiments are classically carried out on uniformly ^{15}N labeled protein so that amide ^{1}H^{15}N bonds form almost isolated spin systems. Data analysis is thus simplified since pure ^{15}N-relaxation mainly accounts for HN flexibility in absence of couplings with neighboring ^{13}C carbons. Dipolar interactions with other protons of the protein are negligible because dipolar interactions are proportional to $1/r^6$. Longitudinal R_1 ^{15}N relaxation, transverse R_2 ^{15}N relaxation and heteronuclear (^{1}H $\rightarrow$ ^{15}N) Overhauser effect were measured.

Relaxation data were analyzed thanks to the spectral density function (Atkinson and Kieffer 2004). The spectral density function is the Fourier transform of the autocorrelation function describing the fluctuations of HN vectors orientations. Qualitatively, it reflects the contribution of HN vectors dynamic fluctuations to relaxation depending on their characteristic time. In the approximation where ^{15}N nuclei relax only via dipole-dipole interactions and chemical shift anisotropy mechanisms, the relaxation rates are related to the spectral density function by the following equation :

$$\begin{pmatrix} R_1 \\ R_2 \\ \sigma \end{pmatrix} = \begin{pmatrix} 0 & 3A+B & A & 0 & 6A \\ \frac{2}{3}(3A+B) & \frac{1}{2}(3A+B) & \frac{1}{2}A & 3A & 3A \\ 0 & 0 & -A & 0 & 6A \end{pmatrix} \times \begin{pmatrix} J(0) \\ J(\omega_N) \\ J(\omega_H - \omega_N) \\ J(\omega_H) \\ J(\omega_H + \omega_N) \end{pmatrix} \tag{6.9}$$

$$avec : A = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_N^2 \gamma_H^2 \hbar^2}{4 r_{NH}^6}$$

$$B = \frac{\Delta\sigma_N^2 B_0^2 \gamma_N^2}{3}$$

where μ_0 is the permeability of the vacuum, γ_N and γ_H are the gyromagnetic ratios of the ^{15}N and ^{1}H nuclei, respectively, $\hbar$ is Planck's constant, r_{NH} is the (^{15}N - ^{1}H) internuclear distance, $\Delta\sigma_N$ is the chemical shift anisotropy of the 15N nucleus, and B_0 is the static magnetic field strength.

It is interesting to note that these three relaxation parameters are sensitive to dynamic fluctuations on different time scales. Heteronuclear nOes σ are only sensitive to fastest fluctuations on timescales around the proton resonance frequency (ν_H = 600 to 950MHz, caracteristic time = 1 to 2ns). Longitudinal relaxation rates R_1 have an important contribution of the spectral density function at the nitrogen resonance frequency (ν_H = 60 to 95MHz) and are sensitive to motions in the 10 to 20ns range. Transverse relaxation rates R_2 is the only parameter to include a $J(0)$ contribution and is therefore sensitive to slower motions and in particular conformational exchange on the micro to millisecond timescale :

$$R_{2obs} = R_2 + R_{ex} \tag{6.10}$$

This exchange term R_{ex} can be evaluated from a linear regression analysis of $R_2 - (R_1/2)$ measured at several fields (Phan et al. 1996). In the fast-exchange regime :

$$\begin{aligned} R_{ex} &= A_{ex} B_0^2 \\ R_2 - \frac{R_1}{2} &= \left(\frac{2\Delta\sigma_N^2 \gamma_N^2}{9} J(0) + A_{ex}\right) B_0^2 + 2A \end{aligned} \tag{6.11}$$

6.6.1 NMR experiments

Relaxation experiments were carried out at ^{1}H frequencies of 600, 800 and 950MHz for CH1 and CH2 for a better description of the spectral density function. For NWASPV2 and CoblAB, relaxation experiments were recorded at 950MHz.

Longitudinal relaxation is the decay constant for the recovery of magnetization along the static magnetic field. Transverse relaxation is the decay constant for the loss of magnetization perpendicular to the static magnetic field.

These relaxation parameters were extracted from pseudo-3D NMR experiments composed of (^{1}H-^{15}N)-HSQC planes with a relaxation delay incrementation (Korzhnev et al. 2001). The general scheme for 2D planes for relaxation experiments is as follows. A preparation step transfers the proton magnetization to the nitrogen to reach a N_z or N_{xy} state for R_1 or R_2 experiments respectively. An incremented delay allows relaxation measurement, followed by ^{15}N chemical shift editing, magnetization transfer to the proton, detection and recycling delay.

Experiments were acquired on an interleaved manner to compensate possible sample evolution over time and differential sample heating depending on relaxation delays.

For the measurement of (^{1}H $\rightarrow$ ^{15}N) heteronuclear nOes, two two-dimensional spectra are recorded. The experiment starts from the ^{15}N magnetization at steady state, with and without prior proton saturation. Experiments were acquired on an interleaved manner.

6.6.2 Data analysis

Spectra were processed with NMRPipe (Delaglio et al. 1995) and intensities were extracted using *nmrDraw* and an in-house-written shell, *anaDyn*.

R_1 and R_2 relaxation rates were obtained by adjustment of the intensity as a function of the relaxation delay to a monoexponential decay. Error on relaxation rate constants was evaluated by Monte Carlo simulations.

$$I(t) = e^{t \times R_i} \tag{6.12}$$

Heteronuclear (^{1}H $\rightarrow$ ^{15}N) nOes were obtained by the ratios of intensities from the spectrum with proton saturation and from the reference spectrum, without proton saturation.

$$\sigma = I_{sat} / I_{ref} \tag{6.13}$$

6.6.3 Analytical models to describe protein dynamics

A global procedure was written for relaxation data analysis using the Octave software. This script adjusts relaxation rates measured at one or several magnetic fields to calculated dynamic parameters using various models for protein dynamics, detailed below. Errors on dynamic parameters were evaluated by Monte Carlo simulations.

Validity of the proposed models is assessed by comparison of the caculated and experimental relaxation parameters and by calculation of the reduced χ^2_{red} for all residues.

$$\chi^2_{red}(i) = \frac{1}{n-p}\left[\left(\frac{R_1^{exp} - R_1^{calc}}{dR_1^{exp}}\right)^2 + \left(\frac{R_2^{exp} - R_2^{calc}}{dR_2^{exp}}\right)^2 + \left(\frac{nOe^{exp} - nOe^{calc}}{dnOe_1^{exp}}\right)^2\right] \tag{6.14}$$

where n is the number of experimental point (here 9 in general, 3 relaxation parameters at 3 magnetic fields), p is the number of adjusted parameters, dR_1^{exp}, dR_2^{exp} and $dnOe_1^{exp}$ and the experimental errors on longitudinal, transverse relaxation rates and heteronuclear nOes respectively.

The CH2 sample was more concentrated than the CH1 sample for measurement of relaxation parameters. As a result the experimental error was lower for CH2, χ^2_{red} were generally higher and were adjusted to allow comparison of the validity of different models for CH1 and CH2.

6.6.3.1 The "model-free" approach

A "model-free" approach was proposed by Lipari and Szabo (1982) for relaxation rate constants interpretation. The protein is treated here as rigid body and internal very flexibility quantified by an order parameter S^2 and a correlation time τ_i is assumed to be decorrelated from the slower overall tumbling quantified by a global correlation time τ_C. The rotational correlation time is the required time for a molecule to rotate one radian. The order parameter quantifies the restriction of the fast HN bond reorientation in a cone of aperture 2θ ($\S^2 = cos(\theta)^2$). The autocorrelation function is therefore the product of the functions for global motions and for local flexibility :

$$G(\tau) = G_{local}(\tau) \times G_{global}(\tau) = S^2 \times e^{-t/\tau_C}(1 - S^2) \times e^{-t/\tau_i} \tag{6.15}$$

The spectral density function obtained by Fourier transform is :

$$J(\omega) = \frac{2}{5}[S^2 \frac{\tau_C}{1 + \omega^2\tau_C^2} + (1 - S^2)\frac{\tau_i}{1 + \omega^2\tau_i^2} \tag{6.16}$$

The spectral density function becomes a simple Lorentz function :

$$J(\omega) = \frac{2}{5}S^2 \frac{\tau_C}{(1 + \omega^2\tau_C^2)} \tag{6.17}$$

These approaches can be completed to take into account possible fast conformational exchange ($R_{ex} = A_{ex}B_0^2$). This description of conformational exchange is only valid is the fast exchange regime when the frequency difference between the two exchanging states($\Delta\omega_N$) is negligible compared to the exchange rate (k_{ex}). The general expression to account for the conformational exchange contribution to relaxation is $R_{ex} = A_{ex} \times B_O^{\alpha}$. With α values of the order of 1 and 0 in the intermediate and slow exchange regimes respectively. It is also possible to add a second order parameter for fast internal motions (S_i^2) (Clore et al. 1990).

The different Lipari & Szabo models used for data adjustment are detailed in table 6.3.

Relaxation data recorded for one magnetic field only providing three constraints per residue for fitting, they are preferentially analyzed with LS1, possibly LS2. Other model-free approaches require relaxation data at several fields to avoid overfitting.

Name	dynamic parameters	spectral density function
LS1	τ_C & S^2	$J(\omega) = \frac{2}{5}\frac{S^2\tau_C}{1+\omega^2\tau_C^2}$
LS2	τ_C, S^2 & τ_i	$J(\omega) = \frac{2}{5}[S^2\frac{\tau_C}{1+\omega^2\tau_C^2} + (1-S^2)\frac{\tau_i}{1+\omega^2\tau_i^2}]$
LS3	τ_C, S^2 & A_{ex}	$J(\omega) = \frac{2}{5}\frac{S^2\tau_C}{1+\omega^2\tau_C^2}$ $R_2 = R_{2i} + A_{ex}B_0^2$
LS4	τ_C, S^2, τ_i & A_{ex}	$J(\omega) = \frac{2}{5}[S^2\frac{\tau_C}{1+\omega^2\tau_C^2} + (1-S^2)\frac{\tau_i}{1+\omega^2\tau_i^2}]$ $R_2 = R_{2i} + A_{ex}B_0^2$
LS5	τ_C, S^2, A_{ex} and α	$J(\omega) = \frac{2}{5}\frac{S^2\tau_C}{1+\omega^2\tau_C^2}$ $R_2 = R_{2i} + A_{ex}B_0^\alpha$
LS6	τ_C, S^2, τ_i & S_i^2	$J(\omega) = \frac{2}{5}[S^2S_i^2\frac{\tau_C}{1+\omega^2\tau_C^2} + (S_i^2-S^2)\frac{\tau_i}{1+\omega^2\tau_i^2}]$

Table 6.3 – *Model-free approaches for the description of protein backbone dynamics*

This approach is classically used to analyze the dynamics of folded proteins with a defined shape and global correlation time τ_C identical for all residues. For the analysis of the relaxation of CH1 and CH2, we used a local model-free approach which consists in adjusting a global τ_C each residue (Alexandrescu and Shortle 1994). These τ_C are associated with reorientation of segment of the peptide chain and this approach if particularly adapted of proteins with both folded regions and dynamic linkers.

6.6.3.2 Distribution of correlation times

For intrinsically disordered proteins, the description of the backbone motions by two dominant motions with associated τ_C and τ_i might be insufficient. These proteins exist as ensemble of rapidly interconverting conformers therefore it might be more relevant to use a distribution of correlation times to describe the ensemble of transitions between conformers (Ochsenbein et al. 2002).

The spectral density function is described as follows :

$$\begin{aligned} J(\omega) &= \int_{-\infty}^{+\infty} F(\tau) J(\omega,\tau)\,\mathrm{d}\tau \\ J(\omega,\tau) &= \frac{2}{5}\frac{\tau}{1+(\omega\tau)^2} \end{aligned} \tag{6.18}$$

The lorentzian distribution is characterized by :

$$\begin{aligned} F(\tau) &= K\frac{\Delta}{\Delta^2+(\tau-\tau_0)^2} \text{ for } 0 \leq \tau \leq \tau_{max} \\ F(\tau) &= 0 \text{ for } \tau > \tau_{max} \\ \int_0^{+\infty} F(\tau)\mathrm{d}\tau &= 1 \end{aligned} \tag{6.19}$$

where τ_0 and Δ are the center and width of the distribution. A maximum correlation time τ_{max} is set at 100ns to ensure convergence of $J(0)$. K is the normalization constant for the distribution.

The spectral density function then becomes :

$$\begin{aligned} J_{LD}(\omega) &= \frac{1}{5D}\left(2\tau_0[1+S\omega^2] - 4K\Delta\omega\tau_0 arctg(\omega\tau_{max}) + K\Delta[1-S\omega^2]ln\frac{\Delta^2+(\tau_{max}-\tau_0)^2}{S(1+\omega^2\tau_{max}^2)}\right) \\ S &= (\Delta^2+\tau_0^2),\ D = 1-2(\Delta^2-\tau_0^2)\omega^2+S^2\omega^4. \end{aligned} \tag{6.20}$$

6.6.3.3 Models derived from polymer theory

Relaxation parameters were also analyzed with a model initially developed for the study of polymers dynamics. This model assumes that the dominant motion of the backbone chain is a crankshaft-like transition (see figure 6.9).

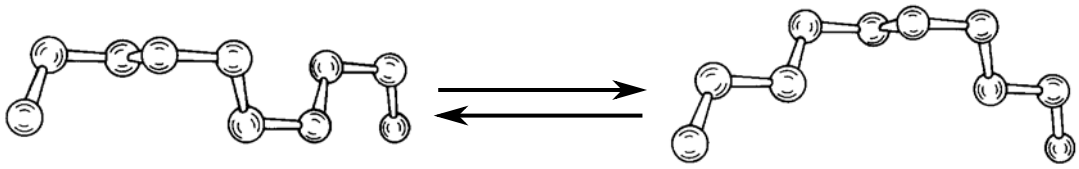


Figure 6.9 – *Nature of the crankshaft transition (Helfand 1971).*

The expressions for the orientation auto-correlation function and spectral density function are :

$$\begin{aligned} G_{HH}(t) &= e^{t/\tau_1}e^{t/\tau_2}I_0(t/\tau_1) \\ J_{HH}(\omega) &= Re(\alpha+i\beta)^{1/2} \\ \alpha &= \tau_2^{-2}+2\tau_1-1\tau_2^{-1}-\omega^2 \\ \beta &= -2\omega(\tau_1^{-1}+\tau_2^{-1}) \end{aligned} \tag{6.21}$$

where I_0 is the zero order modified Bessel function, τ_1 is the characteristic time for the crankshaft-like transition and τ_2 is the characteristic time for the diffusion of the transition along the peptide chain.

DeJean de la Batie et al. (1988a) extended this model to take into account uncorrelated fast HN bond libration in the picosecond range.

$$J_{DLM}(\omega) = AJ_{HH}(\omega) + \frac{2}{5}(1-A)\frac{\tau_f}{1+(\omega\tau_f)^2} \tag{6.22}$$

where A is an order parameter similar to S^2 in model-free analysis.

6.7 Micro- to millisecond dynamics by relaxation dispersion experiments

6.7.1 General introduction

Relaxation dispersion a method of choice for the characterization of conformational exchange on the micro to millisecond timescale when only one major form is observed on NMR spectra (Tolkatchev et al. 2003, Hansen et al. 2009). This is the case in the fast exchange regime on the chemical shifts scale where the observed chemical shifts are weighted averages of the chemical shits of the major and minor states, or in the slow exchange regime when the minor state is only sparsely populated and the minor peaks are not visible on NMR spectra. Relaxation dispersion thus enables characterization of NMR invisible minor states linked to ligand or co-factor binding (Boehr et al. 2006), protein-protein interactions, protein folding and unfolding (Korzhnev et al. 2011) or IDPs folding upon binding (Sugase et al. 2007, Schneider et al. 2015). Applications include studies as elegant as the structure determination of the excited state of a mutated T4 lysozyme using chemical shifts of this minor state determined by relaxation dispersion (Bouvignies et al. 2011).

Figure 6.10 represents the basic principles of the relaxation dispersion experiments (Baldwin and Kay 2009). These experiments allow to measure intensity loss linked to conformational exchange after a Carr-Purcell-Meiboom-Gill (CPMG) delay T_{CPMG} which consists in series of spin echos characterized by a refocalization frequency ν_{CPMG}.

In absence of conformational exchange, 180°pulses of the spin echos refocalize the magnetization whatever the ν_{CP} frequency. Therefore correlation intensities are independent of the ν_{CP} frequency.

For a protein in conformational equilibrium between two states A and B , the two states are alternatively populated over time (figure 6.10 a.). Correlation intensities depend on the population of the minor state, the exchange rate and the interpulse delay. If a transition from A to a state B occurs during a spin echo $t_{CP} = 1/\nu_{CP}$, the precession frequency of an isolated spin probe is modified from ω_A to ω_B. If $\omega_A \neq \omega_B$, that spin probe evolves with different precession frequencies and the magnetization is not completely refocalized leading to intensity loss. This phenomenon is illustrated on figures 6.10 b. and c.. Spin echos during which signal is magnetization is not refocalized are highlighted with grey circles. As the ν_{CP} frequency increases, refocalization, though not perfect, is improved

because spin echos which are not refocused cover shorter periods of times in the T_{CPMG} delay (only a quarter of the T_{CPMG} delay on figure 6.10.), leading to intensity recovery.

Relaxation dispersion profiles are obtained by plotting the effective transverse relaxation R_2^{eff} as a function of the refocusing pulse frequency during the CPMG delay and these curves are analysed to extract exchange parameters : population of the minor state p_E, exchange rate k_{ex} and chemical shift difference between the major and minor state.

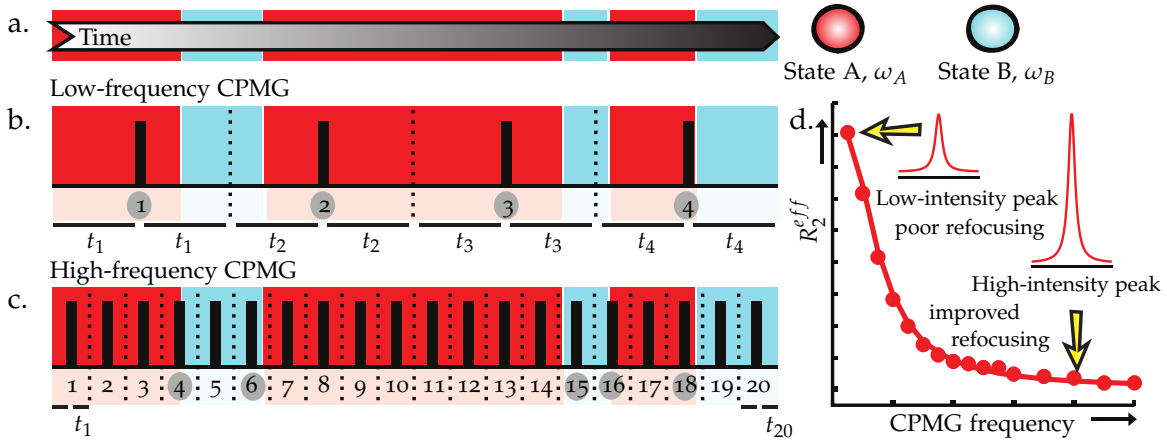


Figure 6.10 – *General principle of relaxation dispersion experiments (Baldwin and Kay 2009) a. A molecule interconverts between states A and B over time. b. and c. Evolution under a CPMG delay with variable ν_{CP} frequencies. Spin echos for which magnetization is not refocalized are shown in grey. d. Typical relaxation dispersion profile displaying the effective transverse relaxation R_2^{eff} as a function of the refocusing pulse frequency during the CPMG delay.*

These experiments are optimal for exchange rates ranging from 200 to 2000Hz. It should also be noted that significant chemical shift differences between the major and minor state are necessary to observe significant dispersion. This method is therefore particularly appropriate to study transitions between folded and unfolded conformers including folding upon binding for IDPs.

6.7.2 NMR experiment

We used ^{15}N-relaxation to probe conformational exchange in free WH2 repeats and folding upon binding actin because they can be recorded on uniformly ^{15}N-labeled samples and do not require sofisticated labeling scheme. Besides HN bonds form isolated spin system so only proton decoupling is required during the CPMG delay.

Relaxation dispersion experiments are pseudo 3D composed of (^{1}H-^{15}N)-HSQC plane with a third dimension corresponding to the ν_{CPMG} frequency incrementation.

A pulse sequence with compensation of nitrogen relaxation (Loria et al. 1999) and proton decoupling (Hansen et al. 2008) was implemented on the spectrometers by Jean-Pierre Placial. The summarized magnetization pathway is the following :

Proteins	Buffer	Temperature	magnetic field	T_{CPMG}
CH1	G-buffer, pH=7	278K	950MHz	100ms
CH2	G-buffer, pH=7	278K	950MHz	100ms
NWASPV2	G-buffer, pH=7	278K	950MHz	50 & 100ms
CoblAB	G-buffer, pH=7	278K	950MHz	40 & 80ms

Table 6.4 – *List of recorded relaxation dispersion experiments on free WH2 repeats*

$$^{1}H_z \xrightarrow{\text{INEPT}} {}^{15}N_xH_z \xrightarrow[T_{CPMG}/2]{\text{CPMG}} N_xH_z \rightarrow N_x \xrightarrow[T_{CPMG}/2]{\text{CPMG}} {}^{15}N_x \xrightarrow{\text{t1/RINEPT/WATERGATE}} {}^{1}H_{x/y}$$

From the amide proton, magnetization is transferred to the nitrogen via an INEPT block. The CPMG include relaxation of the antiphase coherence, refocalization and relaxation of the inphase coherence. This compensated scheme (Loria et al. 1999) allows equal contributions of inphase and antiphase transverse relaxation. The CPMG block is followd by ^{15}N chemical shifts encoding, return INEPT and WATERGATE water suppression.

In addition, a continuous wave proton decoupling is applied during the CPMG delay. To avoid spin-echo disruption by proton decoupling, the decoupling frequency has to be a multiple of two times ν_{CP} (Vallurupalli et al. 2007). So the decoupling pulses length and power are recalculated for every spin echo frequency ν_{CP}. Pulse length and pulse power lists are generated by a tcl script written by Jean-Pierre Placial. Inputs are the duration and power of the 90°proton pulse, mean decoupling frequency, total CPMG delay T_{CPMG} and spin echo frequency ν_{CP} list.

Relaxation dispersion experiments were recorded on free WH2 repeats (CH1, CH2, NWASV2 and CoblAB) at 278K, in G-buffer at pH=7, at the ^{1}H frequency of 950MHz.

It is interesting to work on spectrometers operating at high frequency for this kind of experiments because exchange contribution to the transverse relaxation R_{ex} is proportional to B_0^{α} with α ranging from 0 to 2 for slow to fast exchange regime. Thus profiles are susceptible to show more important dispersion at higher magnetic field allowing better characterization of intermediate to fast exchange processes.

Relaxation dispersion experiments were also recorded on CH1 and CoblAB in presence of actin at 298K, in G-buffer at pH=7.

6.7.3 Data analysis

Pseudo-3D NMR spectra are processed with *nmrPipe* and intensities were extracted using an in-house program *anaDyn*.

For each residue and each ν_{CPMG} frequency, effective transverse relaxation rates are calculated assuming a monoexponential decay :

$$
\begin{aligned}
I^{\nu_{CPMG}}(i) &= I_0(i) \times exp(-R_{2eff}^{\nu_{CPMG}}(i) \times T_{CPMG}) \\
R_{2eff}^{\nu_{CPMG}}(i) &= \frac{1}{T_{CPMG}} \frac{I_{ref}(i)}{I_{\nu_{CPMG}}(i)}
\end{aligned}
\tag{6.23}
$$

Exchange parameters (population of the minor state p_B, exchange rate k_{ex} and chemical shift difference between the free and bound states) are adjusted using an *Octave* script written by Jean-Pierre Placial (Placial 2013).

Theoretical R_{2eff}^{calc} for a set of exchange parameters are calculated by solving the Block-McConnell equations numerically for an isolated ^{15}N spin is exchange between two states A and B :

$$
\mathrm{A} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{B}
$$

$$
A = \begin{bmatrix} \frac{-i\Delta\omega}{2} - R_2 - k_{on} & k_{off} \\ k_{on} & \frac{-i\Delta\omega}{2} - R_2 - k_{off} \end{bmatrix}
\tag{6.24}
$$

Adjustment of the exchange parameters is achieved by minimization of the χ^2 for individual residues or groups of residues in a global conformational exchange model :

$$
\chi^2 = \sum_{i,\nu CPMG} \frac{\left(R_{2eff}^{exp}(i, \nu_{CPMG}) - R_{2eff}^{calc}(i, \nu_{CPMG}\right)^2}{\sigma^2}
\tag{6.25}
$$

6.8 Millisecond dynamics by CEST experiments

6.8.1 Introduction to saturation transfer experiments

Saturation transfer can be used to detect invisible minor species in solution that can be saturated by off-resonance radiofrequency (RF) radiation without irradiating the major species, this saturation is then transferred to the visible state or molecule by chemical exchange. By looking at intensity profiles of major NMR visible species as a function of RF irradiation offset, one can obtain information on minor NMR invisible species in chemical exchange with the major observed resonance.

In small molecule saturation transfer difference (STD), broad resonances of a macromolecule are saturated, saturation is transferred to a bound ligand by cross-relaxation and finally to the detected unbound ligand by chemical exchange (Meyer and Peters 2003, Fielding 2007) enabling compounds screening for binding activity, epitope mapping for identification of groups in direct contact with the protein and determination of binding constants for interacting molecules. These techniques are also used in imaging where molecules with resonance frequencies distinct from water are saturated and saturation is transferred to water by slow chemical exchange to provide additional information or improve contrast of images (Zhou and van Zijl 2006).

More recently, dark-state exchange saturation transfer (DEST Fawzi et al. (2011; 2012)) and chemical exchange saturation transfer (CEST Vallurupalli et al. (2012)) NMR spectroscopy were applied to detection of minor states in protein macromolecules. DEST can be used to characterize slow exchange with large assemblies or aggregates and was applied to study of $A\beta$ protofibrils formation (Fawzi et al. 2011), IAPP (Brender et al. 2015) and poyubiquitin (Morimoto et al. 2015) aggregation and interaction of $A\beta$ with the GroEL chaperone (Libich et al. 2013). Relatively large B_1 frequencies are used for saturation (more that 100Hz) and the intensity profiles are highly sensitive to transverse relaxation of the minor state.

On the other hand, CEST experiments allow to probe slow conformational exchange linked to protein folding/unfolding or protein/protein and protein/ligand interactions. Applications of ^{15}N-CEST experiments include identification of an off-pathway intermediate in an acyl carrier protein domain unfolding (Lim et al. 2014), study of cold an hot denaturation of a disulfide rich PAF protein (Fizil et al. 2015), analysis of slow conformational dynamics of small GTPases complexed with GTP and GTP analogues (Long et al. 2013), characterization of the interaction of the intrinsically disordered regulatory domain of MKK7 with c-Jun N-terminal kinase (JNK1 Kragelj et al. (2015)). Lower B_1 field frequencies are used (5 to 50 Hz) which allows better resolution for the determination of chemical shifts of the invisible minor state : intensity losses are observed when the B_1 field is applied at the frequency offsets of the two exchanging states. Effect of various exchange parameters on ^{15}N-CEST profiles are highlighted on figure 6.11 for exchange between two states with chemical shifts of ω_G = 121.8ppm and ω_E = 119ppm.

Carbon CO, C_α and methyl CEST experiments were also developed however additional carbon-carbon and nitrogen-carbon scalar couplings within uniformly ^{15}N-^{13}C-labeled protein leads to sensitivity loss and more complex evolution matrix for CEST profiles analysis.

6.8.2 Pulse sequence implementation

The ^{15}N-CEST experiment is a pseudo-3D experiment comprising a series of (^{1}H-^{15}N)-HSQC planes with a relaxation delay during which a weak RF field is applied at various ^{15}N offset frequencies (Vallurupalli et al. 2012). The summarized magnetization pathway is the following :

$$^1\mathrm{H}_z \xrightarrow[\text{INEPT}]{\text{refocused}} {}^{15}\mathrm{N}_z \xrightarrow[\text{CEST block}]{T_{EX}} {}^{15}\mathrm{N}_z \rightarrow {}^{15}\mathrm{N}_x \xrightarrow{\text{t1/SE-RINEPT}} {}^1\mathrm{H}_{x/y}$$

The pulse sequence was adapted from a bruker (^{1}H-^{15}N)-HSQC via double inept transfer for T1 relaxation measurement with sensitivity improvement (hsqct1etf3gpsi3d) The T1 relaxation delay was replaced by a CEST block (see figure) : a weak ^{15}N B_1 field irradiation with a frequency ranging from 5 to 50Hz and ^{1}H decoupling to reduce the ^{1}H-^{15}N spin system to an isolated ^{15}N spin. The RF B_1 field offsets are listed in a frequency list (FQ3LIST) were offsets are given in Hz par rapport to the ^{15}N offset for the spectrum. ^{1}H

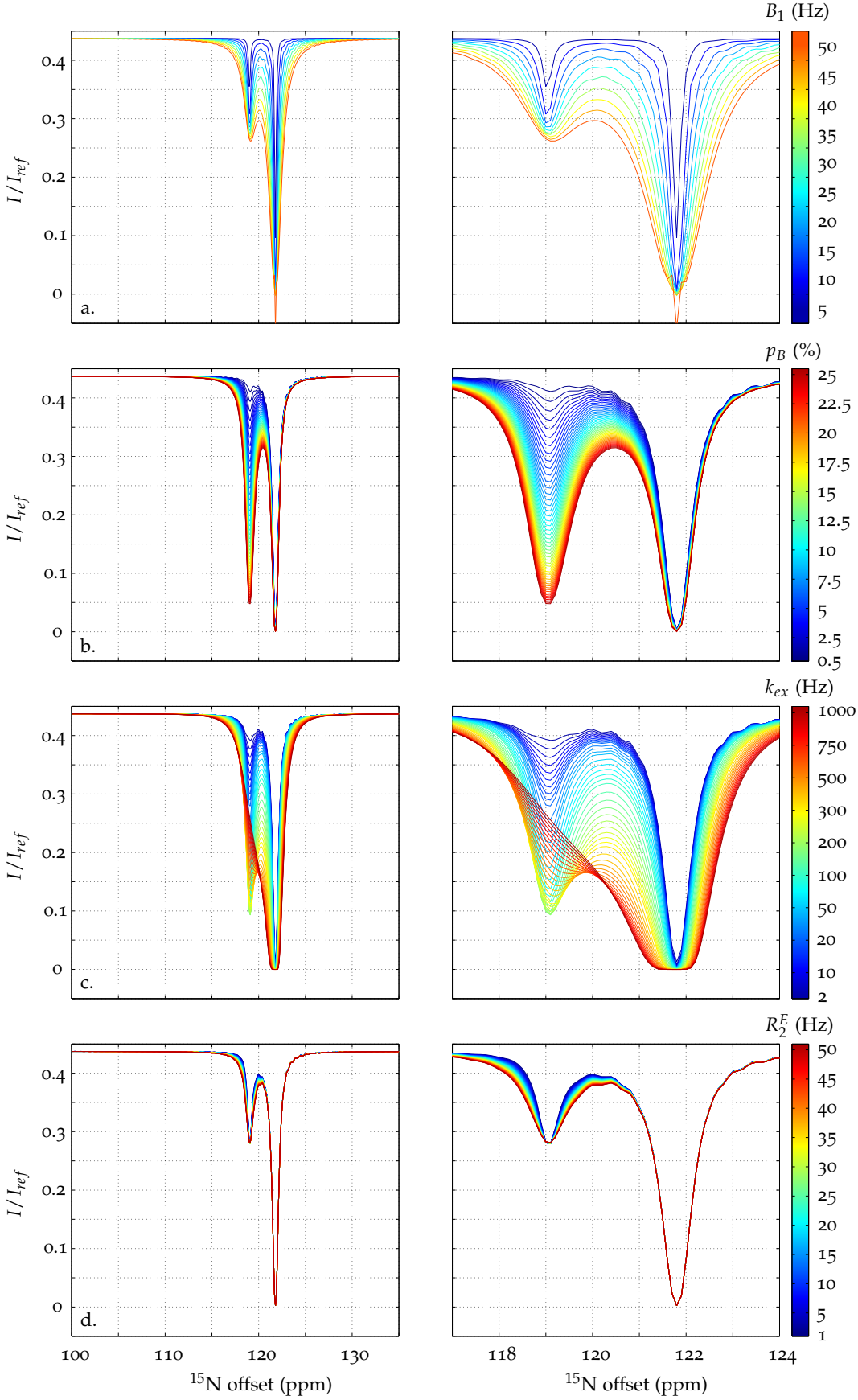


Figure 6.11 – *Effect of the B_1 field frequency (a.), the population of the minor state p_E (b.), the exchange rate between G and E states k_{ex} (c.) and the transverse relaxation of the minor state R_2^E (d.) values on CEST profiles. B_1 inhomogeneity is 10% for all simulations. Default B_1, p_E, k_{ex} and R_2^E values are 25Hz, 5%, 20Hz and 25Hz respectively.*

decoupling is achieved by a composite 180°pulse ($90_x240_y90_x$) (Levitt 1982) to minimize side bands artifacts (Vallurupalli et al. 2012) in the intensity profiles. HSQC planes corresponding to different B_1 field offsets were acquired in an interleaved manner to avoid intensity differences between CEST planes due to sample evolution over time.

The full pulse program is shown in appendix.

6.8.3 Data analysis

Spectra were processed with NMRPipe (Delaglio et al. 1995) and intensities were extracted using an in-house-written program, *anaDyn*.

Intensities curves as a function of ^{15}N offset frequencies were then adjusted to extract possible exchange parameters using the *Chemex* program developed by Guillaume Bouvignies (Vallurupalli et al. 2012). Theoretical intensities for a set of parameters were calculated using the Bloch-McConnell equations which describes the evolution during a CEST block for a single spin -1/2 particle in exchange between two states G and E :

$$\frac{d}{dt}\begin{pmatrix} E/2 \\ I_x^G \\ I_y^G \\ I_z^G \\ I_x^E \\ I_y^E \\ I_z^E \end{pmatrix} = M \times \begin{pmatrix} E/2 \\ I_x^G \\ I_y^G \\ I_z^G \\ I_x^E \\ I_y^E \\ I_z^E \end{pmatrix}$$

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^G - k_{GE} & -\omega_G & \omega_1 & k_{EG} & 0 & 0 \\ 0 & \omega_G & -R_2^G - k_{GE} & 0 & 0 & k_{EG} & 0 \\ 2R_1^G I_{eq}^G & -\omega_1 & 0 & -R_1^G - k_{GE} & 0 & 0 & k_{EG} \\ 0 & k_{GE} & 0 & 0 & -R_2^E - k_{EG} & -\omega_E & \omega_1 \\ 0 & 0 & k_{GE} & 0 & \omega_E & -R_2^E - k_{EG} & 0 \\ 2R_1^E I_{eq}^E & 0 & 0 & k_{GE} & \omega_1 & 0 & -R_1^E - k_{EG} \end{pmatrix}$$

where k_{GE} and k_{EG} are the exchange rates between the G and E states. I_{eq}^G and I_{eq}^E are the equilibrium populations of the G and E states. ω_G and ω_E are the offsets (in rad/s) of the weak ^{15}N irradiation field from states G and E. ω_G is not adjusted but obtained from the major NMR visible state. ω_1 is the pulsation of the weak RF field ($\omega_1 = \gamma_{^{15}N} B_1$). And R_1^G, R_2^G, R_1^E and R_2^E are the ^{15}N relaxation rates for the G and E states. R_1^E is assumed equal to R_1^G because its value has little influence on CEST profiles. R_1^G is determined using a reference experiment with $T_{EX} = 0$. For B_1 field irradiation at ^{15}N offsets away from G and E state :

$$I(T_{EX}) = I_0 exp(-R_1^G T_{EX}) \Leftrightarrow R_1^G = \frac{1}{T_{EX}} ln(\frac{I_0}{I})$$

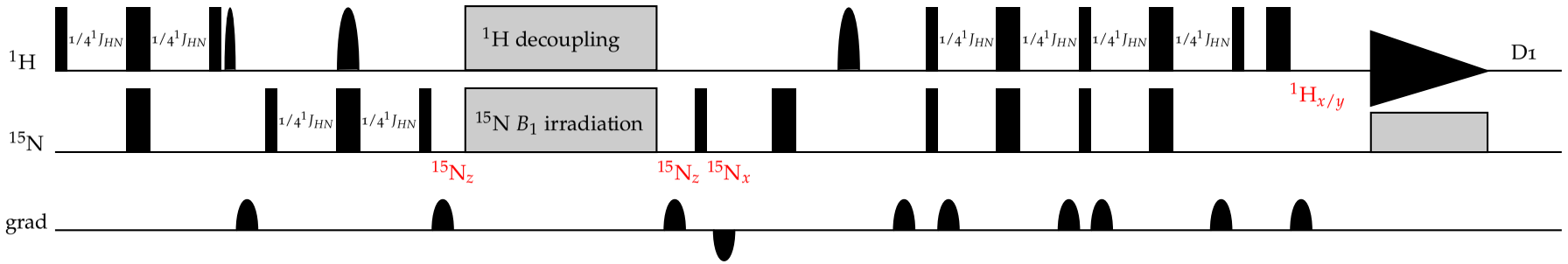


Figure 6.12 – *Pulse sequence for the ^{15}N-CEST experiment. 90°and 180°hard pulses are shown as narrow and wide black bars. Soft pulses and gradients have rounded shapes. The ^{1}H and ^{15}N transmitters are positioned on water and at the center of the amide nitrogen region (between 118 and 122ppm depending on the protein) respectively except during the CEST block when they are moved at the center of the amide protons region and at the desired ^{15}N offset (determined by the FQ3list).*

Therefore CEST profiles depend on 6 adjusted parameters : p_E, $k_e x$, ω_E, R_1^G, R_2^G, and R_2^E. For a global exchange, p_E and $k_e x$ are identical for all fitted residues and do not depend on the spectrometer and B_1 field frequencies. R_1^G, R_2^G, and R_2^E are residue specific, depend on the spectrometer frequency but not on B_1 frequency. To improve the fit, ^{15}N-CEST experiments were recorded for several B_1 field frequencies, typically 13 and 27Hz.

6.8.4 Validation on a model protein

To validate the pulse sequence implemented on the spectrometer, CEST experiments were carried out on a model protein : the domain 1 of human annexin 1 (A1D1). The dynamic behavior of this protein was extensively characterized in the laboratory ((Gallopin 2007, Gallopin et al. 2004)). Human annexin 1 is a 346-residue protein comprising four domains of identical topology. Only A1D1 (81 residues from Gly33 to Lys113) forms an autonomous folding unit in solution, similar to its structure in the full protein (Gao et al. 1999).

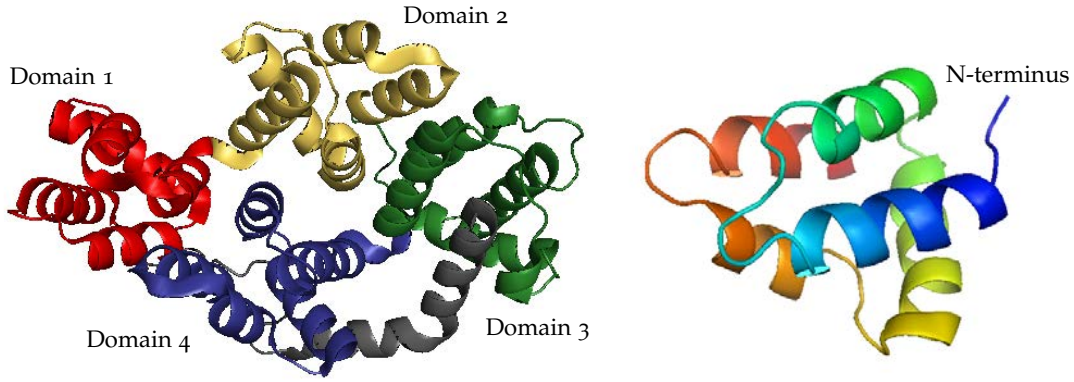


Figure 6.13 – *Structure of full human annexin 1 (left, PDB : 1HM6) and of the isolated domain 1 (right, PDB : 1BO9)*

The domain stability depends on the pH and at pH=3, the protein is in slow equilibrium between folded and unfolded states.

^{15}N-relaxation dispersion and ^{15}N-CEST experiments were recorded on the same uniformly ^{15}N-labeled sample at 298K, pH=3 and at ^{1}H frequency of 800MHz.

Relaxation dispersion profiles were recorded using a CPMG relaxation delay of 40ms and ν_{CPMG} values of 0, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 650, 750, 850 and 1000Hz.

Spectra were analyzed as described in the previous section. The 49 R_{ex} profiles showing significant dispersion were fitted with a two-states global exchange exchange model to extract exchange parameters : the minor state population p_E=1.1% the exchange rate t_{ex}=1.40ms or ν_{ex}=714Hz and the chemical shift differences between the major folded state and the minor unfolded state (see figure 6.14). The global χ^2_{glob} value was 1976. Chemical

exchange is distributed all along the sequence in agreements with a two-state unfolding of A1D1.

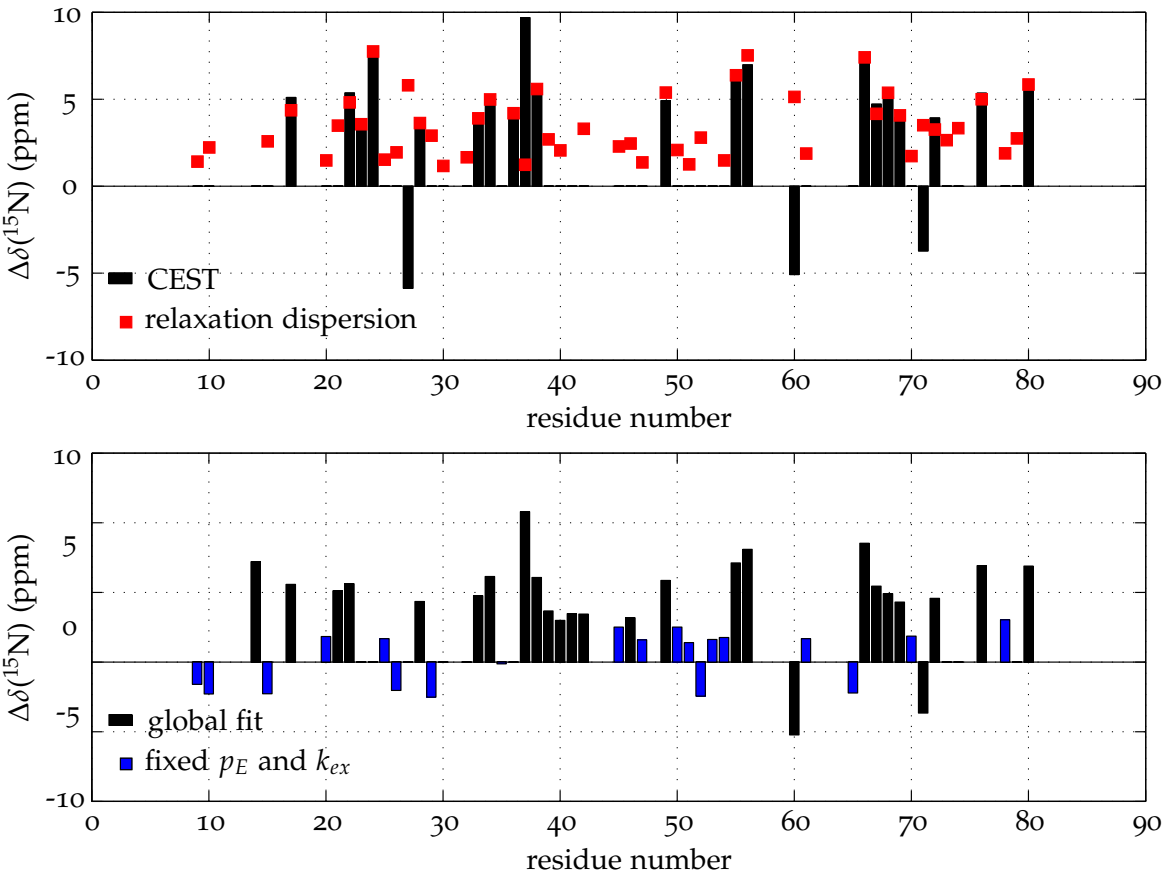


Figure 6.14 – *On top, fitted ^{15}N chemical shift differences between folded and unfolded A1D1 at 298K and pH=3 determined by relaxation dispersion and CEST experiments. On bottom, fitted ^{15}N chemical shift differences between folded and unfolded A1D1 at 278K and pH=3 determined by CEST experiments. Residues used for the determination of a global p_E and k_{ex} are in black.*

The ^{15}N-CEST profiles were acquired using a delay of 400ms for the CEST block and ^{15}N offsets for the B_1 field ranging from 102.7ppm to 133.8ppm with intervals of 0.5ppm (40Hz). Significant exchange was only observed on 23 residues (in red on figure). Global data fitting on these 23 residues enabled to extract the minor population $p_E = 0.92 \pm 0.02\%$ and the exchange rate $k_{ex} = 814 \pm 26$Hz for a two states exchange model as well as the chemical shifts of the minor state (see figure 6.14). These values are in agreement with those determined using relaxation dispersion experiments.

Chemical shift differences for the minor state obtained by fitting of the relaxation dispersion curves are absolute values. The sign of this difference can be determined using an additional set of 2D experiments (Skrynnikov et al. 2002). On the other hand, the sign of the chemical shift difference is given by the CEST profile.

CEST experiments are optimal to characterize exchange ranging from 20 to 200Hz. For faster exchange rates, peaks linewidth increases as shown on figure 6.11c. leading to a unique asymmetric peak rather than two well defined peaks on the CEST profiles. It

Proteins	Buffer	Temperature	magnetic field	T_{EX}	B_1
CH2	G-buffer	278K	950MHz	400ms	25Hz
CH2 + 2.5% actine	G-buffer	278K	950MHz	400ms	25Hz
CH2 + 2.5% actine	G-buffer	298K	950MHz	400ms	25Hz
CH1 + 5% actine	G-buffer	298K	800MHz	400ms	25Hz 12.5Hz
CH1 + 5% actine	80mM KCl	298K	800MHz	400ms	25Hz 12.5Hz
CH1 + 5% actine	100mM KCl	298K	800MHz	400ms	25Hz 12.5Hz
CH1 + 5% actine	125mM KCl	298K	800MHz	400ms	25Hz 12.5Hz
CH1 + 5% actine	150mM KCl	298K	800MHz	400ms	25Hz 12.5Hz
CoblAB	G-buffer	278K	800MHz	400ms	25Hz
CoblAB + 10% actine	G-buffer	298K	800MHz	400ms	25Hz 12.5Hz

Table 6.5 – *List of recorded ^{15}N-CEST experiments*

explains why chemical exchange was only observed on 23 CEST profiles for which the chemical shift difference was above 3.5ppm.

^{15}N-CEST experiments were also recorded on the same sample at 278K to slow down the exchange rate. CEST profiles displayed two obvious intensity dips for 42 residues and a single asymetric dip for 3 residues. 26 residues for which the CEST profiles showed to well resolved peaks (in orange on figure **??**) were fitted to a two states exchange model to determine the minor population $p_E = 1.50 \pm 0.04\%$ and the exchange rate $k_{ex} = 63 \pm 3$Hz. The other curves were fitted using these fixed values for p_E and k_{ex}.

6.8.5 CEST experiments on WH2 repeats

^{15}N-CEST experiments where recorded on CH1, CH2 and CoblAB in absence of actin to probe millisecond conformational exchange processes in the free protein and in presence of actin to follow the folding upon binding process. Actin in CEST experiments is recombinant 5C-actin produced using the baculovirus/Sf9 system (see section 4.2).

III

Results

Recombinant actin

7

We wanted to produce recombinant actin for several reasons. Actin purification from rabbit muscles is constraining in terms of logistics and we are limited to skeletal α-actin. Recombinant expression enables one to express any actin isoform from any species, as well as engineered mutants for structural studies. Although isotopic labeling in insect cells is still on development, actin labeling for direct observation by NMR will enable to address unanswered questions linked to actin dynamics and conformational plasticity. We did not have the competencies nor the infrastructure to express proteins in insect cells at the ICSN at the time this project started so this work was a collaboration with Stéphane Pêtres and Christine Girard-Blanc (Protein production in eucaryotic cells platform, Pasteur Institute). This chapter presents the characterization of wild-type and mutant recombinant actins and isotopic labeling strategies in insect cells for direct actin observation by NMR.

7.1 Production of recombinant actin

7.1.1 Preparation of recombinant actin

Both wild-type drosophila 5C-actin and the non-polymerizable AP-actin mutant were expressed and purified with high yield using the baculovirus/Sf9 system (see section 4.2). Expression level was similar for the two proteins as evaluated by SDS-PAGE and western-blot (anti-Strep antibody). Typical yields were 10mg per liter of culture after the Streptrag purification for the two constructs. Further gel-filtration purification was necessary to separate actin monomers from another form of lower elution volume, probably unfolded actin or actin dimers. Final yield after gel-filtration was 5mg per liter of culture.

Cleavage of the Strep-tag led to actin precipitation. Therefore, both the wild-type and AP actins have a 42-residue-long extension at the N terminus, composed of the double Strep-tag and PreScission recognition site. The presence of this N-terminal extension did not abolish the polymerizing properties of 5C actin. The sequences of rabbit actin and 5C-actin display a 93% sequence identity with 1 deletion in drosophila 5C-actin and 23 substitutions (shown on figure 7.2). Among these distinct residues, 14 have strongly similar properties (scoring > 0.5 in the Gonnet PAM 250 matrix), 7 have weakly similar properties (scoring ≤ 0.5 in the Gonnet PAM 250 matrix) and 2 have different properties. Therefore interaction with WH2 repeats is not expected to be dramatically modified.

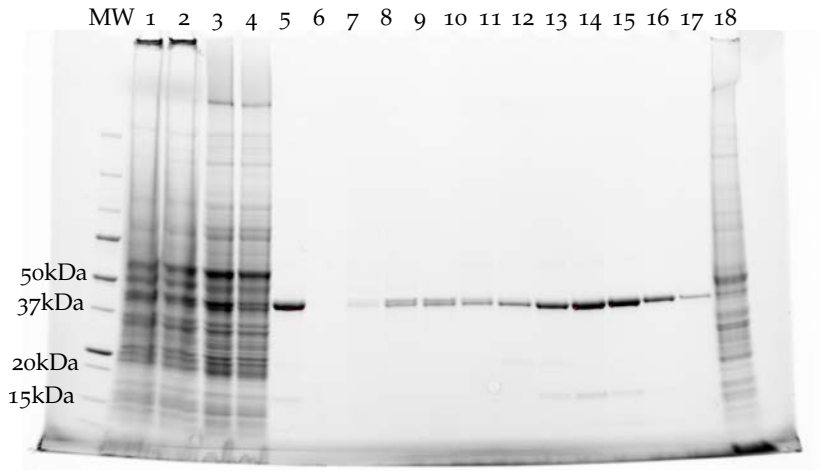
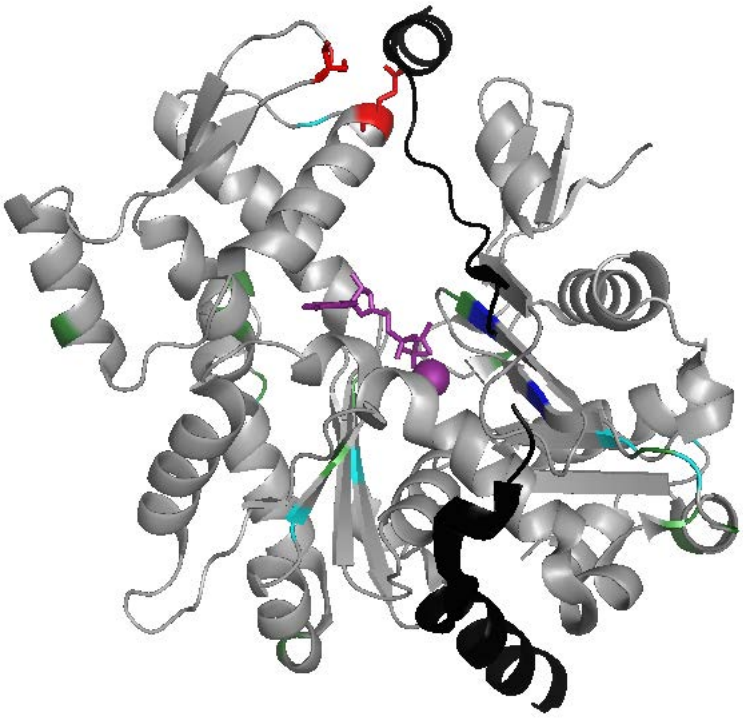


Figure 7.1 – *SDS-5-15% polyacrylamide gel - analysis of 5C-actin purification. Lane 1 is the cell lysate. Lane 2 is the supernatant after ultracentrifugation. Lane 3 is the supernatant after overnight dialysis. Lane 4 is the efluent after Streptrap column binding. Line 5 shows the actin pool after Streptrap purification. Lanes 6 to 17 are fractions of the gel-filtration purification. Lanes 12 to 17 show the purified actin monomers. A minor contaminant of 17kDa, possibly cofilin as described in Iwasa et al. (2008) can be observed on lanes 13 to 15. Lane 18 shows the pellet after overnight dialysis.*



Rabbit actin/5C-actin	
E2-	M176L
D3E	V201T
T5V	N225Q
T6A	T260A
C10V	I267L
L16M	S271A
V17C	A272C
I76V	I287V
T103V	N297T
V129T	M299L
T160S	T358S
N162T	A365S

Figure 7.2 – *Crystal structure of drosophila AP-actin in the Ca-ATP-bound state (PDB : 2HF4) in grey (ATP and calcium are shown as purple sticks and sphere) superimposed with a model of actin-bound CH2 (PDB : 3U9D and 1T44) in black. Mutated residues in AP actin (E204 and K243) are shown in red. Sequences differences between rabbit actin and drosophila 5C actin are highlighted in green for substitutions with residues of strongly similar properties, in cyan for substitutions with residues of weakly similar properties, in blue for substitutions with residues of different properties and in yellow for the deletion.*

7.1.2 Thermal Shift Assay

Stability of the recombinant actins was evaluated by thermal shift assay. These experiments were realized by Eric Jacquet and Naima Nhiri. High background fluorescence signal at low temperature is probably due to the exposed hydrophobic cleft between subdomains 1 and 3. In G-buffer, the two constructs have an identical melting temperature of 62°C. Therefore the double point mutation (A204E and P243K) introduced in AP-actin does not change the stability of the protein. Besides, this value is in good agreement with previous results obtained on ATP-Ca-bound α-skeletal rabbit actin in G-buffer by DNase I-inhibition assay (60.4°C, Schuler et al. (2000)) or by differential scanning calorimetry (61°C, Pivovarova et al. (2013))

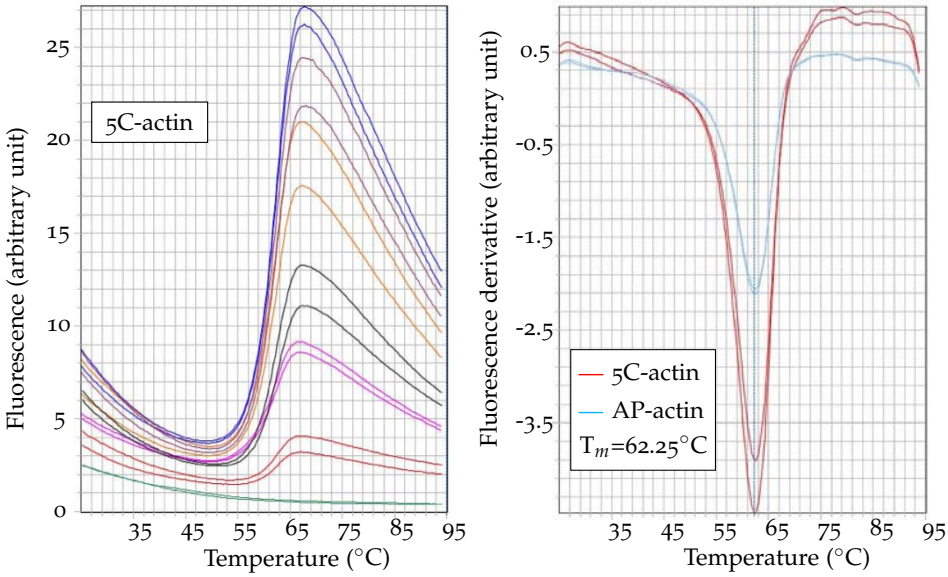


Figure 7.3 – *Stability of recombinant 5C- and AP-actin evaluated by thermal shift assay. a. Raw fluorescence melting curves of 5C-actin at various concentrations. b. First derivative of the fluorescence curves, the minimum at 62.25°C corresponds to the melting temperature for 5C- and AP-actin*

7.2 Interaction with CH2

To determine whether recombinant 5C and AP actins can be used for further structural studies to investigate interaction with WH2 repeats, we compared binding of a model single WH2 domain, CH2, to these two recombinant actins and to tissue purified α-skeletal rabbit actin.

CH2 is a sequestering single WH2 domain, chimera of human thymosin β4 and of the first domain of drosophila Ciboulot. Interaction of CH2 with rabbit α-skeletal actin was investigated by combined X-ray diffraction, NMR, SAXS and biochemical studies (Didry et al. 2012). The free protein is mainly disordered and folds upon binding actin. An amphipathic N-terminal helix binds the hydrophobic cleft between subdomains 1 and 3 of actin, followed by an extended central region and a C-terminal helix that caps the

pointed end of actin between the subdomains 2 and 4 (see figure 7.2). This sequestering complex is stable for 5 days in the NMR spectrometer and is formed with low risks of actin polymerization.

7.2.1 Microscale thermophoresis

The dissociation constant K_D for the (CH2-actin) complex was previously derived from AEDANS fluorescence titration curves using AEDANS-labeled rabbit α-skeletal G-actin (K_D of 1nM).

Here we used microscale thermophoresis to determine the dissociation constants associated to the interaction of CH2 with α-skeletal rabbit actin and recombinant 5C and AP-actin. The sensitivity of the method as well as the use of capillaries greatly reduces the amount of protein needed. The same actin concentration (50nM), MST power and LED power were used for the 3 constructs to allow better comparison of the results. Raw normalized fluorescence timetraces acquired during MST experiments are represented on figure 7.4.

Initial fluorescence of rabbit actin at low CH2 concentration is lower than that of recombinant drosophila actins (figure 7.5.a) in agreement with the lower labeling efficiency (50% for rabbit actin and 70% for recombinant actins).

Initial fluorescence increases by 86%, 40% and 36% for rabbit, 5C and AP actin respectively upon addition of CH2 as shown on figure 7.5.a which indicates a modification of the fluorophore's environment upon CH2 addition. As explained in the material and methods, rabbit actin has one main reactive cystein (Cys374) whereas recombinant drosophila actin has two additional accessible cysteins (Cys2 and Cys272). Therefore labeling of recombinant actins might be divided among these three sites. Cys374 is located near the hydrophobic cleft between the subdomains 1 and 3 which is the binding site for amphipathic helices of WH2 repeats. Cys2 in the subdomain 1 and Cys272 in the subdomain 3 are more distant from WH2 repeats binding sites (see figure 5.3). This is consistent with the smaller fluorescence increase observed for recombinant drosophila actins where only part of the fluorophores undergo environment modification upon CH2 binding.

For the three actin constructs, a simple two-states binding mode is sufficient to describe the initial fluorescence variation. The dissociation constants K_D are obtained by adjustment of the fluorescence curves as described in the material and methods and the adjusted parameters are summarized in table 7.1. Calculated initial fluorescence and percentage in fluorescence change are plotted on figures 7.5.a and b.

CH2 binds rabbit actin and recombinant 5C-actin with a similar affinity whereas affinity is about 4 to 5 times lower for recombinant AP-actin.

The difference in normalized fluorescence is plotted on figure for analysis of thermophoresis with temperature jump. These profiles show a more complex biphasic behavior. Indeed, the initial fluorescence is sensitive to the fluorophore's environment whereas microscale thermophoresis also probes changes in the size, charge and hydration shell of

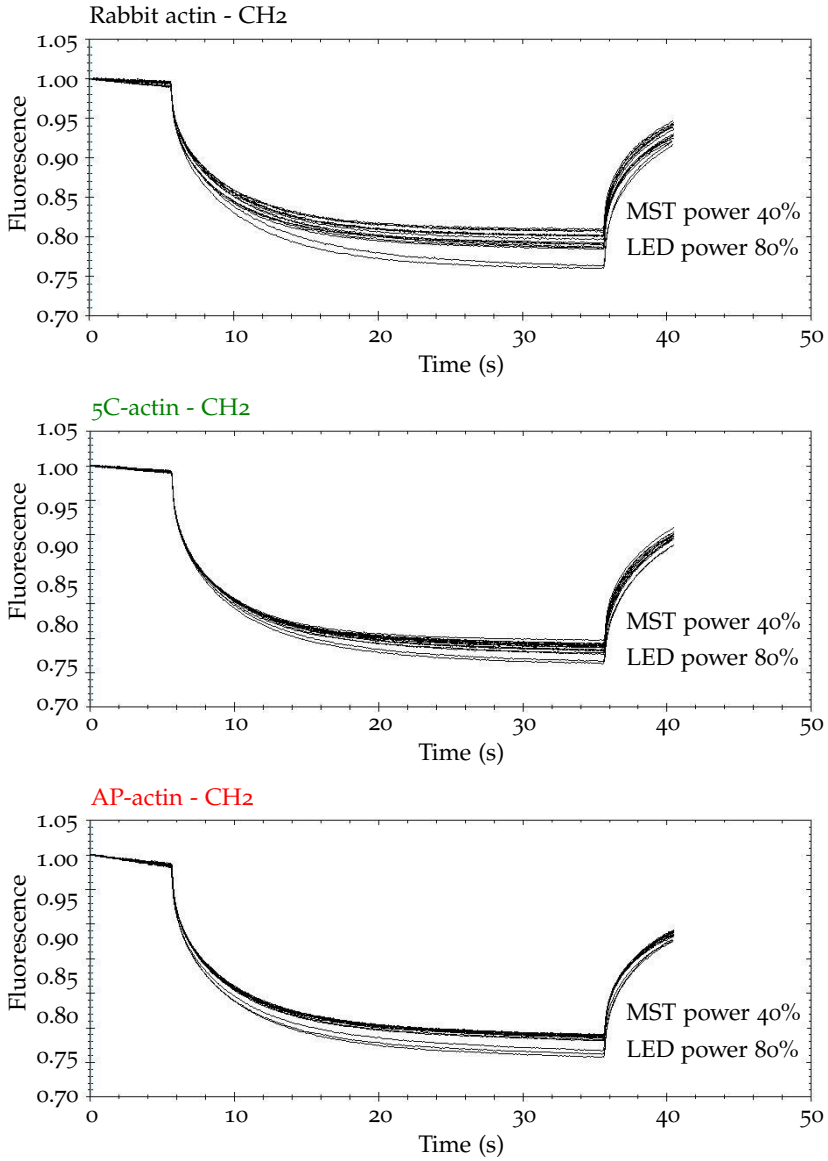


Figure 7.4 – *Raw normalized fluorescence timetraces acquired during MST experiments. The different curves on each graph correspond to different CH2 concentrations. The initial fluorescence signal is normalized to 1. The fluorescence is initially acquired for 6 seconds. The temperature gradient is applied for 30 seconds and turned off while the fluorescence signal is acquired for 5 seconds.*

Protein	$F_{unbound}$ (AU)	F_{bound} (AU)	K_D (nM)
Rabbit actin	656 ± 18	1223 ± 24	104 ± 29
5C-actin	1036 ± 14	1449 ± 12	124 ± 32
AP-actin	1097 ± 14	1487 ± 17	499 ± 119

Table 7.1 – *Adjusted parameters from fluorescence curves fits. $F_{unbound}$ (F_{bound}) is the initial fluorescence of a 50nM of labeled actin solution in MST buffer in absence of CH2 (fully bound to CH2).*

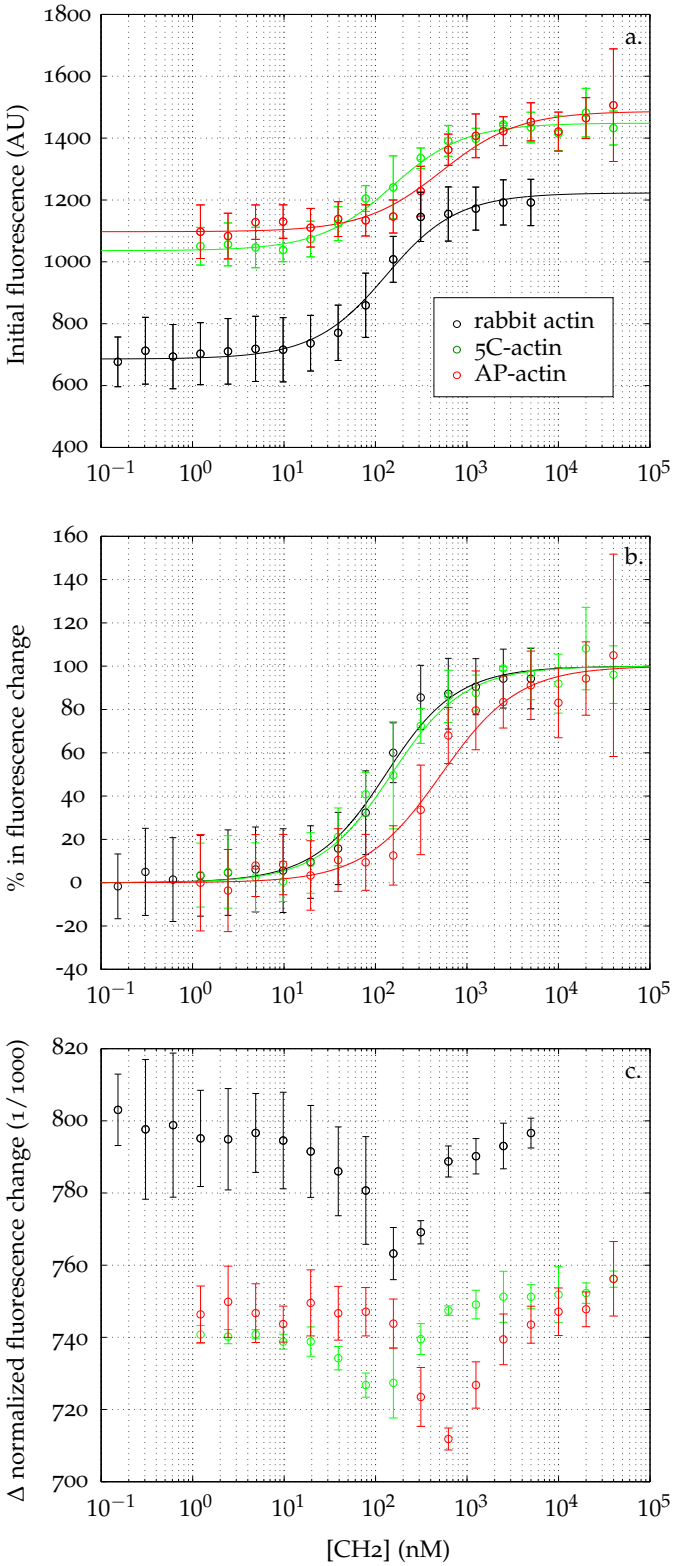


Figure 7.5 – *a. Initial fluorescence signal during MST experiments. b. Percentage in the fluorescence change upon addition of CH2. c. MST signal including temperature jump.*

the labeled protein. Opposite thermophoretic effects may add up making data hard to interpret. Interestingly, the minima in MST profiles seem to correspond to the inflection points of the fluorescence curves.

For rabbit α-skeletal actin, the dissociation constant derived from MST experiments ($K_D = 104 \pm 29nM$) is two orders of magnitude higher that the one previously determined using AEDANS fluorescence titration curves ($K_D = 1nM$). Several factors may explain this difference :

1. Tween20 was added to the G buffer used for MST experiments. Although the concentration is very low (0.05%), this detergent might reduce binding affinity.

2. There was no change in AEDANS-actin fluorescence upon CH2 addition, in G-buffer. Therefore affinity was evaluated by titration curves of AEDANS-actin by Tβ4 in the presence of competitive amounts of CH2. This indirect method may also introduce additional bias in the K_D determination.

3. Actin was labeled with different fluorophores : AEDANS in one case and Nanotemper NT-495 fluorescent dye in the other. The company does not give the structure of the NT-495 fluorescent dye, so it is not possible to discuss this point in detail but different fluorophore sizes and chemical properties may influence binding affinity, especially since the reactive cysteine 374 is close to the binding site of the amphipathic helix.

7.2.2 NMR experiments

Interactions of recombinant 5C and AP actins with CH2 were also analyzed by NMR spectroscopy.

The ^{1}H-^{15}N HSQC spectrum of CH2 bound to actin from rabbit muscle was previously assigned by François-Xavier Cantrelle. The ^{1}H-^{15}N HSQC spectrum of CH2 bound to recombinant 5C actin (figure 7.6) is very similar and the assignment was directly transferred.

The presence of the N-terminal tag does not disturb the interaction of the amphipathic helix of CH2 with actin. Chemical shift differences are observed between CH2 bound to rabbit actin or 5C actin (blue curve on figure 7.6). They may be due to the sequence differences between rabbit actin and 5C actin. For instance, the 3 different amino acids in the beta sheet of the subdomain 3 may cause the shifts above 0.07ppm for residues 34 to 37 of CH2. However, these differences are small compared to the chemical shift differences between free CH2 and rabbit actin bound CH2 (average of 0.04ppm and 0.57ppm respectively). We therefore conclude that CH2 interacts similarly with rabbit actin and recombinant 5C actin.

The ^{1}H-^{15}N HSQC spectrum of CH2 bound to recombinant AP-actin is represented

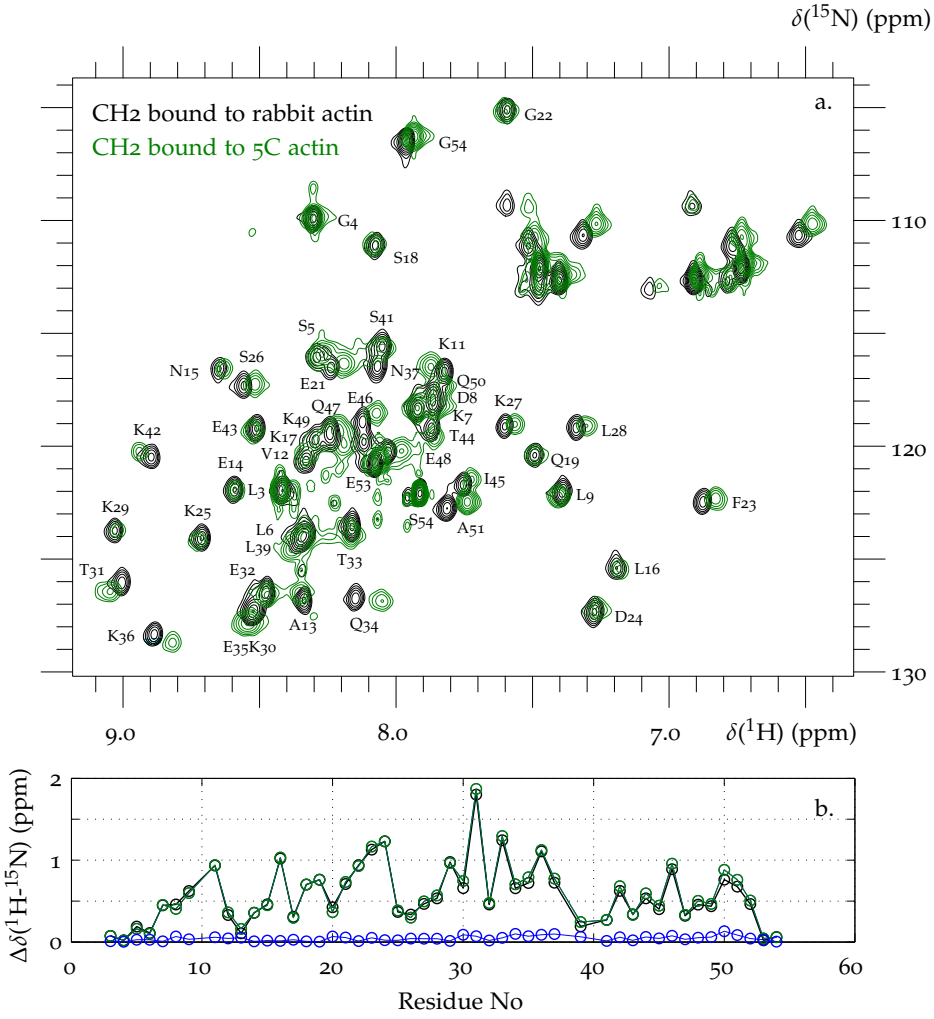


Figure 7.6 – *a.* ^{1}H-^{15}N-*HSQC spectra of CH2 in complex with rabbit actin (black) and 5C- actin (green). b. Normalized chemical shift difference between CH2 bound to rabbit actin and free CH2 (black), between CH2 bound to 5C-actin and free CH2 (green), between CH2 bound to rabbit actin and CH2 bound to 5C-actin (blue)*

on figure 7.7 and is significantly different than the spectrum of CH2 bound to 5C-actin. Two groups of peaks can be identified. The first group displays an important spectral dispersion in the proton dimension consistent with a folded region. The peaks were assigned by direct transfer from the spectrum of CH2 bound to rabbit actin and correspond to the residues of the N-terminal half of CH2 from Lys7 to Lys30 which contains the N-terminal helix and the first half of the LKKT motif. The residues Glu14 and Asn15, at the N-terminus of the amphipathic helix, have split peaks on the HSQC indicating that two states are in slow exchange at the chemical shift scale. The second group displays a small spectral dispersion in the proton dimension, peaks are narrower and more intense, consistent with a disordered region. Besides, these peaks superimpose with the spectrum of free CH2 (see figure 7.7) so they could also be assigned by direct transfer of the assignment of

free CH2. They correspond to the residues of the C-terminal half of CH2 from Asn37 to Ser54.

We recorded 15N transverse relaxation experiments for free CH2 and for CH2 bound to AP-actin. The resulting ralaxation rate constants are represented on figure 7.8 for residues 37 to 54. Peaks of N-terminal fragment residues were too weak to determine R_2 accurately. For the (CH2:AP-actin) complex, R_2 relaxation decreases through the C-terminal region to reach values comparable to the free CH2 for the last residues. This is another indication that flexibility increases as one moves away from the region bound to actin. The lysine 42 has a higher R_2 relaxation than its neighbor residues which might be due to chemical exchange on the micro to millisecond timescale.

The mutations introduced in AP-actin are both located in the interaction region of the C-terminal helix of CH2. The proline 243, mutated into a lysine in AP-actin, is close to the lysine 42 of CH2 which might interact with aspartate 244 of actin. The alanine 204 of rabbit and 5C actin is close to the isoleucine 45 of CH2 and the mutation into a glutamate may also destabilize the contacts between the C-terminal helix of CH2 and AP-actin.

Therefore NMR data indicates that CH2 does interacts with AP-actin via its N- terminal amphipathic helix. However, the mutations introduced in the subdomain 4 of actin to make it non polymerizable also disturb the interaction of the C-terminal helix of CH2 with AP-actin which is released and remains disordered in the (CH2 : AP-actin) complex.

7.3 Isotopic labeling of proteins overexpressed in insect cells

Isotopic labeling in insect cells was done in collaboration with François Bontems and Annalisa Meola who performed all experiments on S2 cells and who did most of the optimization work on S2 cells because smaller volumes were required thanks to high production and purification yields of DIII-EFF and DIII-AFF. Four constructs were expressed in S2 cells : DIII-EFF-1 (Arg409-Leu609) and DIII-AFF-1 (Ser388-Ala485, Mw 15 kDa) designed to correspond to the domains III of the *Caenorhabditis elegans* EFF-1 and AFF-1 fusion proteins, DIII- HaV (Met283-Asp415, Mw 18kDa) designed to comprise the domain III of the Hantaan virus Gc fusion protein based on the prediction for its ortholog in Andes hantavirus, DIII-RVFV (Ala336-Ser457, Mw 17 kDa) corresponding to domain III of the Rift valley fever virus Gc protein. These constructs were fused with a signal peptide for protein secretion and proteins were purified from the culture supernatant. AP-actin was expressed using the baculovirus/Sf9 system. This work was published in the Journal of Structural Biology (Meola et al. 2014)

7.3.1 Motivations and general strategy

Conventional protein isotopic labeling schemes for NMR applications rely on overexpression in *Echerichia coli*. This bacterium can grow on minimal media where ammonium chloride and glucose are the sole nitrogen and carbon sources respectively and can be

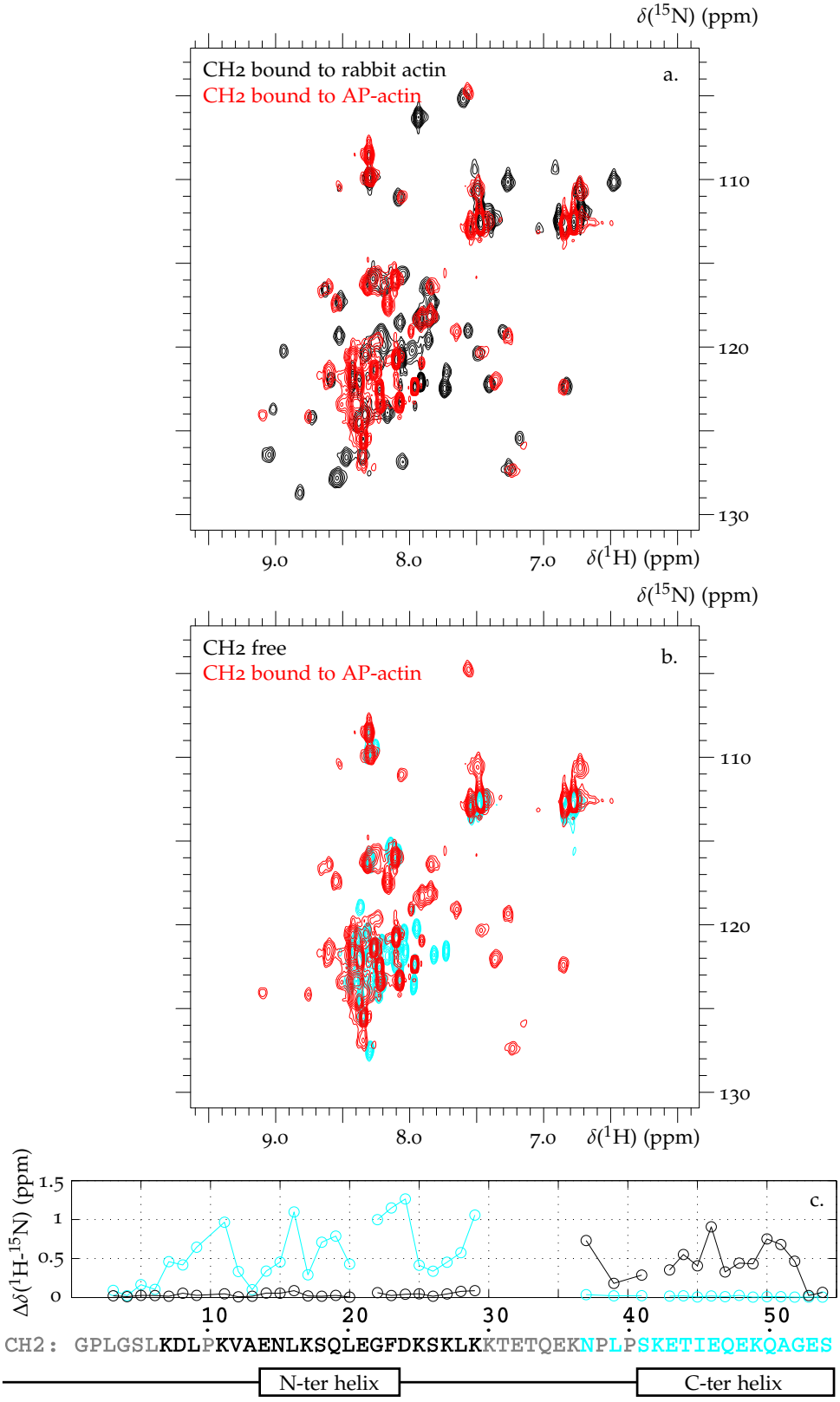


Figure 7.7 – *a.*^{1}*H-*^{15}*N-HSQC spectra of CH2 in complex with rabbit actin (black) and AP- actin (red). b.* ^{1}*H-*^{15}*N-HSQC spectra of CH2 in complex with AP-actin (red) and of free CH2 (cyan). c. Normalized chemical shift difference between CH2 bound to AP-actin and CH2 bound to rabbit actin (black), between CH2 bound to AP-actin and free CH2 (cyan). The sequence of CH2 is displayed below.*

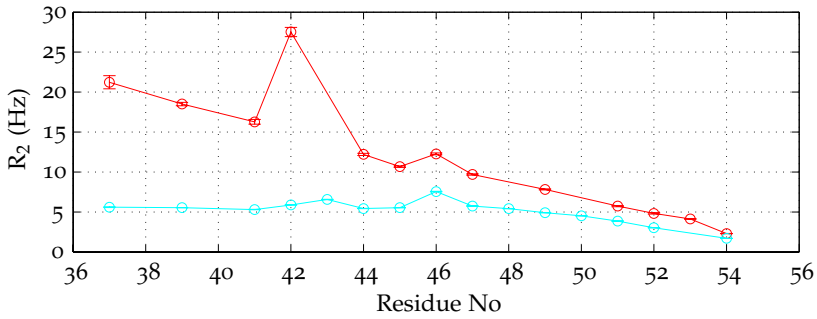


Figure 7.8 – ^{15}N *transverse relaxation of free CH2 (cyan) and CH2 bound to AP-actin (red) for residues 37 to 54.*

uniformly labeled when using $^{15}NH_4Cl$ and ^{13}C-glucose. Besides, *E. coli* tolerates the replacement of water by deuterated water enabling deuteration of overexpressed proteins. The metabolic pathways of *E. coli* are well described and various methods were developed for non-uniform labeling such as specific labeling using labeled amino acids, semi-uniform labeling using [1,3-^{13}C]-glycerol and [2-^{13}C]-glycerol (LeMaster and Kushlan 1996, Castellani et al. 2002) introduction of ^{1}H-^{13}C labeled methyl groups in a deuterated context(Gardner and Kay 1997, Gans et al. 2010).

However many eucaryotic and viral proteins cannot be overexpressed in *E. coli* because they need specific chaperones for proper folding, like actin which requires the chaperonin complex CCT, or because they need post-translational modifications to be functional. Insect cells provide a good compromise for production of these kind of proteins because they are easier to handle than mammalian cells and often provide relevant post-translational modification. Insect cells only grow on complex media most of which are synthetic proprietary media of unreleased composition.

Complex media for uniform isotopic labeling of protein overexpressed in insect cells do exist but are highly expressive : about 10000 euros for 1L of *Cambridge Isotope Laboratories* ^{15}N,^{13}C-Bioexpress 2000 compared to about 400 euros for 1L of (^{15}N,^{13}C)-M9 minimal medium suitable for overexpression in *E. coli*. These commercial labeling media were only scarcely used for uniform ^{15}N- and ^{15}N-^{13}C-labeling using the baculovirus/Sf9 system to characterize the dynamic interactions of the Abelson kinase domain with several clinical inhibitors (Strauss et al. 2005, Vajpai et al. 2008) and for *in-Sf9 cell* NMR of the *Streptococcus* protein G B1 domain (Hamatsu et al. 2013), and using mammalian HEK293 (Human Embryonic Kidney) and A549 (human epithelial) cells (Sastry et al. 2011) for expression of HIV gp120 glycoprotein.

Complex media depleted in some or all amino acids are commercially available. Supplemented with the relevant amino acids, they were used for specific ^{15}N and/or ^{13}C labeling (Gossert et al. 2011, Nygaard et al. 2013). However applications are limited to very specific cases and assignment of the resonances is not straightforward. and may be time-consuming. For instance, Nygaard et al. (2013) assigned $^{13}CH_3\epsilon$-Met resonances by mutagenesis which requires the preparation of viruses for all the mutants.

This strategy is not appropriate for uniform labeling because of the high price of some amino acids. An alternative is to add an amino acid mixture obtained from hydrolysate or autolysate of organisms grown on a labeled minimal medium instead of using purified amino acids. Previous attempts are summarized in table 7.2.

Cells	Source of amino acids	Yield*	Reference
Mammalian CHO	algea hydrolysate + Cys + Gln	-	Hansen et al. (1992)
Insect Sf9	algea hydrolysate + Cys + Glu + Gln + Trp + NH_4Cl	60%	Walton et al. (2006)
Mammalian CHO & HEK293	yeast autolysate + algae hydrolysate	40-60% (CHO) 60-80% (HEK293)	Egorova-Zachernyuk et al. (2011)

Table 7.2 – *Protein labeling in insect or mammalian cells. * Protein production yield in labeled medium compared to expression in standard rich medium*

Further attention should be focused on the composition of the amino acid mixture, glutamine metabolization in glutamate and alanine biosynthesis from pyruvate.

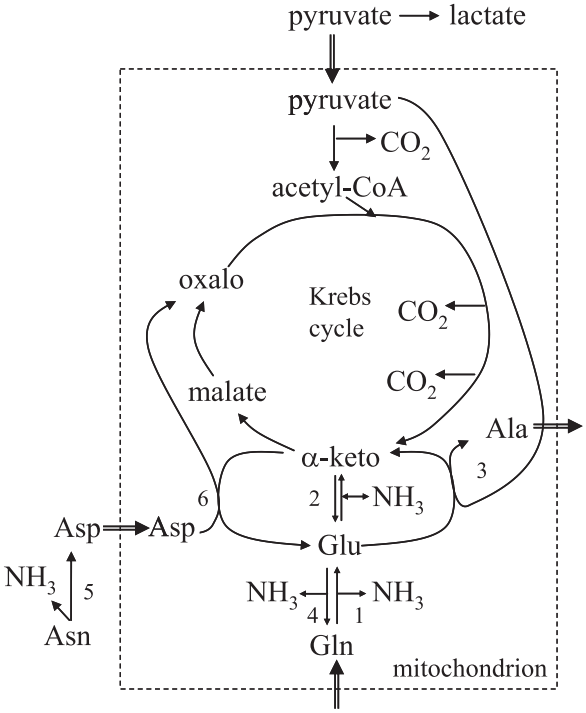


Figure 7.9 – *Gln/Glu and Asn/Asp metabolism in insect cells (Drugmand et al. 2012). Involved enzymes are : 1, glutaminase; 2, glutamate dehydrogenase; 3, alanine aminotransferase; 4, glutamine synthetase; 5, asparaginase; 6, aspartate aminotransferase.*

Acid hydrolysis to obtain algae extract leads to degradation of fragile amino acids (glutamine, asparagine, cysteine, and tryptophan) therefore media using these hydrolysate have to be supplemented with these labeled amino acids increasing the medium cost. On the other hand, yeast extracts obtained by yeast digestion by their own proteases contain all amino acids.

Another issue is the high level of glutamine in insect cells culture media required for protein synthesis but also for energetic metabolism. Direct addition of ^{15}N-glutamine highly increases the medium cost. Enzymatic synthesis of ^{15}N-glutamine from ^{15}N-glutamate is possible but adds an additional time-consuming step. Ohman et al. (1995) showed it is possible to induced glutamine synthesis from glutamate in Sf9 insect cell by addition of NH_4Cl to the culture medium. Walton et al. (2006) took advantage of this metabolic switch to reduce the quantity of glutamine added to the labeling medium.

The main pathway for alanine biosynthesis is NH_3 transfer to pyrurate. Supplementation of the medium with $^{15}NH_4Cl$ should be should be sufficient to avoid isotopic dilution in nitrogen labeling. However control of ^{13}C-labeling requires addition of ^{13}C-labeling requires addition of labeled glucose as the expense of an important increase of the medium cost.

In conclusion, although it could help to characterize proteins of paramount interest in solution, production of isotopically labeled samples overexpressed in insect cells remains confined to a few examples because of the lack of simple, robust and cost-effective methods.

7.3.2 Medium formulation and protocol optimization

The conditions for the design of the labeling medium were the following :

- The medium should be significantly cheaper than the available commercial media
- There should be as little steps as possible for the medium preparation at the bench to ensure robustness, simplicity and reproducibility.
- Protein production yield should be of the same order of magnitude as in the unlabeled medium.
- The medium should be suitable for transient expression using the baculovirus/Sf9 system, as well as expression induced in stable cell line (S2 cells).
- The medium should be adaptable to various labeling schemes.

We decided to use a rich medium depleted in all amino acids (ESF-Δaa) supplemented with yeast autolysate rather than algae hydrolysate as the main amino acid source because yeast autolysates obtain yeast digestion by their own proteases contain all amino acids. ^{15}N-labeled yeast autolysate was provided by *Bio Springer*, and contains 50% of free amino acids with an over-representation of alanine, glycine, leucine, serine and threonine and an under-representation of methionine and glutamine compared to rich ESF921 culture medium. Three yeast extract concentrations of 20, 10 and 5g/L later referred as YA_{20}-ESF, YA_{10}-ESF and YA_5-ESF respectively were tested.

DIII-EFF was overexpressed by S2 cells in YA_{20}-ESF complemented with 2mM of glutamine and in YA_{10}-ESF complemented with 5mM of NH_4Cl. Purification yield were sim-

ilar indicating that it was possible to lower the yeast extract concentration to 10g/L and to replace glutamine by NH_4Cl.

We first thought cells had to be adapted to the labeling medium containing yeast extract in particular because amino acid distributions in YA_{10}-ESF and ESF921 are very different. Cells were first adapted from Insect-Xpress (which was the standard medium in the laboratory) to ESF921 and from ESF921 to YA_{10}-ESF. This adaptation step was time-consuming (2-3 weeks) and although both S2 and Sf9 cells were able to grow on YA_{10}-ESF medium for a few weeks, we had no idea if they would be stable on the long run. DIII-AFF and DIII-HaV were overexpressed by S2 cells in ESF921 rich medium and after transfer in YA_{10}-ESF and YA_5-ESF with no adaptation step just before induction. Cell viability was similar showing it is possible to remove the adaptation phase from ESF921 to YA_{10}-ESF and that it is also possible to use 5 g/L of yeast autolysate without compromising cell viability. Our protocol was thereafter simplified by growing S2 and Sf9 cells and transferring when in YA_{10}-ESF or YA_5-ESF just before induction of S2 cells or after baculovirus infection of Sf9 cells.

Production efficiency of S2 and Sf9 cells in YA_{10}-ESF, YA_5-ESF and Insect-Xpress were compared. Actin overexpression in Sf9 cells was first evaluated qualitatively by western-blot. No significant decrease in actin expression was observed between unlabeled Insect-Xpress and labeled YA-ESF media. Protein expression in S2 and Sf9 cells was then analyzed quantitatively by comparison of yield after Streptrap purification (see table 7.3). Similar production yields were obtained in Insect-Xpress and YA_{10}-ESF media whereas it was significantly lower in YA_5-ESF (decrease of 40% for AP-actin and 35% for DIII-HaV compared to YA_{10}-ESF). Therefore a concentration of 10g/L of yeast extract was used for further experiments.

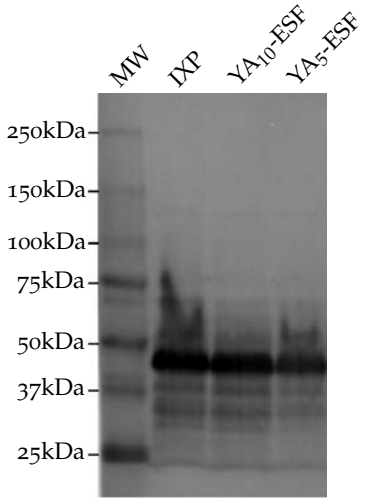


Figure 7.10 – *Qualitative analysis of actin overexpression in Insect-Xpress (IXP), YA_{10}-ESF and YA_5-ESF media by anti-Strep western-blot*

Cells	Protein	IXP	YA_{10}-ESF	YA_5-ESF	mL for 20nmol
Sf9	AP-actin	6 [3-10]	10 [9-10]	6	83
S2	DIII-HaV	22 [18-28]	17[13-19]	11 [7-16]	22
S2	DIII-RVFV	10	12		28
S2	DIII-AFF1	4	4		75

Table 7.3 – *Protein expression yields obtained measured after strep-tag purification in mg/L of culture with S2 and Sf9 cells with Insect-Xpress and with YA_{10}-ESF or YA_5-ESF without adaptation. For AP-actin and DIII-HaV, productions were realized over a period of several months, the mean, minimal and maximal values (measured after strep-tag purification) are indicated in mg/L of culture. The last column indicates the volume necessary to produce about 20 nmol of the proteins, this value corresponding to an NMR sample of 100 µM in 200 µL (sample volume in a 3 mm NMR tube).*

Production of DIII-HaV in Insect-Xpress lead to an heterogeneous product with the expected 18kDa protein and degradation products of lower molecular weight. In YA_{10}-ESF, protein degradation is strongly enhanced and only the shortest 15kDa form is recovered. Mass spectrometry experiments showed that the shortest form lacked 25 amino acids compared to the full form. Comparison of the ^{1}H-^{15}N HSQC spectra of the full and degraded DIII-Hav showed that the missing amino acids belonged to a disordered region at the N-terminus (data not shown). This increase in protein degradation may be caused either by proteases released by insect cells or by proteases already present in YA_{10}-ESF medium. To test the latter hypothesis, purified DIII-HaV expressed in Insect-Xpress was incubated for 48 hours in Insect-Xpress, YA_{10}-ESF and YA_{10}-ESF treated for protease inactivation (tYA_{10}-ESF). Degradation was only observed in the standard YA_{10}-ESF medium. Therefore *Bio Springer* added a protease inactivation step to the preparation of yeast autolysate for protein expression. Protease inactivation is crucial for expression of secreted proteins containing disordered or flexible and accessible regions.

Final conditions for ^{15}N-labeling in insect cells :
The labeling medium is an amino acid deficient medium (ESF-Δaa) supplemented with 10g/L of ^{15}N-labeled yeast extract treated for proteases inactivation and 5mM $^{15}NH_4Cl$. Insect cells are grown in unlabeled rich medium (here Insect-Xpress). Cells are centrifuged just before expression transferred in the labeling medium after infection by recombinant baculovirus for Sf9 cells or before induction for S2 cells. Protein expression is then carried out according to standard protocols.

7.3.3 Results on actin labeling

Isotopic labeling was performed on AP-actin rather than 5C-actin to enable collection of NMR spectra in absence of binding partner without risking polymerization. ^{15}N-labeled AP-actin was successfully overexpressed in tYA_5-ESF and tYA_{10}-ESF. The sofast-^{1}H-^{15}N HMQC spectrum of ^{15}N-labeled actin collected at 950MHz and 298K is represented on

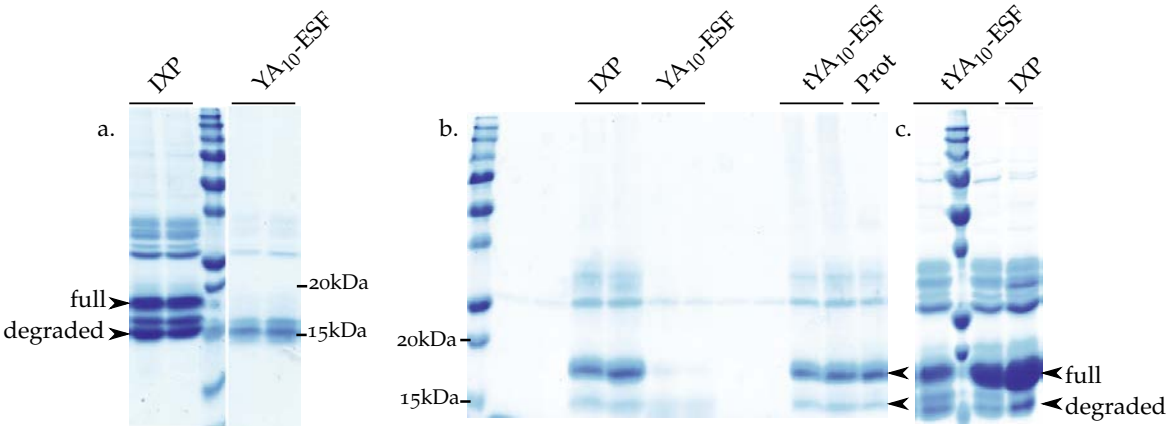


Figure 7.11 – *Comparison of the effect of the Insect-Xpress, YA_{10}-ESF and YA_{10}-ESF treated for protease inactivation (tYA_{10}-ESF), on the integrity of DIII- HaV that was shown to possess an N-terminal unstructured extension. a. SDS–PAGE analysis of the protein expressed in Insect-Xpress (IXP) and YA_{10}-ESF. In Insect-Xpress, the major band at 18 kDa corresponds to the full protein and the two others to partially degraded forms (at the level of the N-terminus). Only the most degraded form of the protein is found in the case of the protein expressed in YA_{10}-ESF. b. SDS–PAGE analysis of the protein expressed in Insect-Xpress, purified and incubated in the three culture media during 48 hours. c. SDS–PAGE analysis of the protein expressed in Insect-Xpress and tYA_{10}-ESF.*

figure 7.12. The spectrum is characteristic of a folded protein with an important spectral dispersion in the proton dimension.

A total of 396 peaks are expected for backbone amide resonances (418 amino acids - 21 prolines - 1 N-terminus). In addition, 48 peaks for asparagine and glutamine side chains, 18 peaks for arginine side chains and 6 peaks for tryptophane side chains are expected. About 300 peaks were identified on the sofast-^{1}H-^{15}N HMQC spectrum 12 of which in the arginine side chain region ($\delta(^{15}\mathrm{N}) \simeq$ 80-90ppm) out of 18 expected peaks. Missing peaks may be due to superimposition especially in the tag regions which contains repeated motifs and line broadening.

		AP-actin	N-ter tag
backbone	residues	376	42
	prolines	18	3
		358 peaks	38 peaks
side chains	asparagine	9 (18 peaks)	0
	glutamine	12 (24 peaks)	3 (6 peaks)
	arginine	18	0
	tryptophane	4	2
		64 peaks	8 peaks

Table 7.4 – *Details of the peaks expected on the sofast-^{1}H-^{15}N HMQC spectrum of ^{15}N-labeled AP-actin*

^{15}N isotopic incorporation was estimated by recording two-dimensional ^{1}H experi-

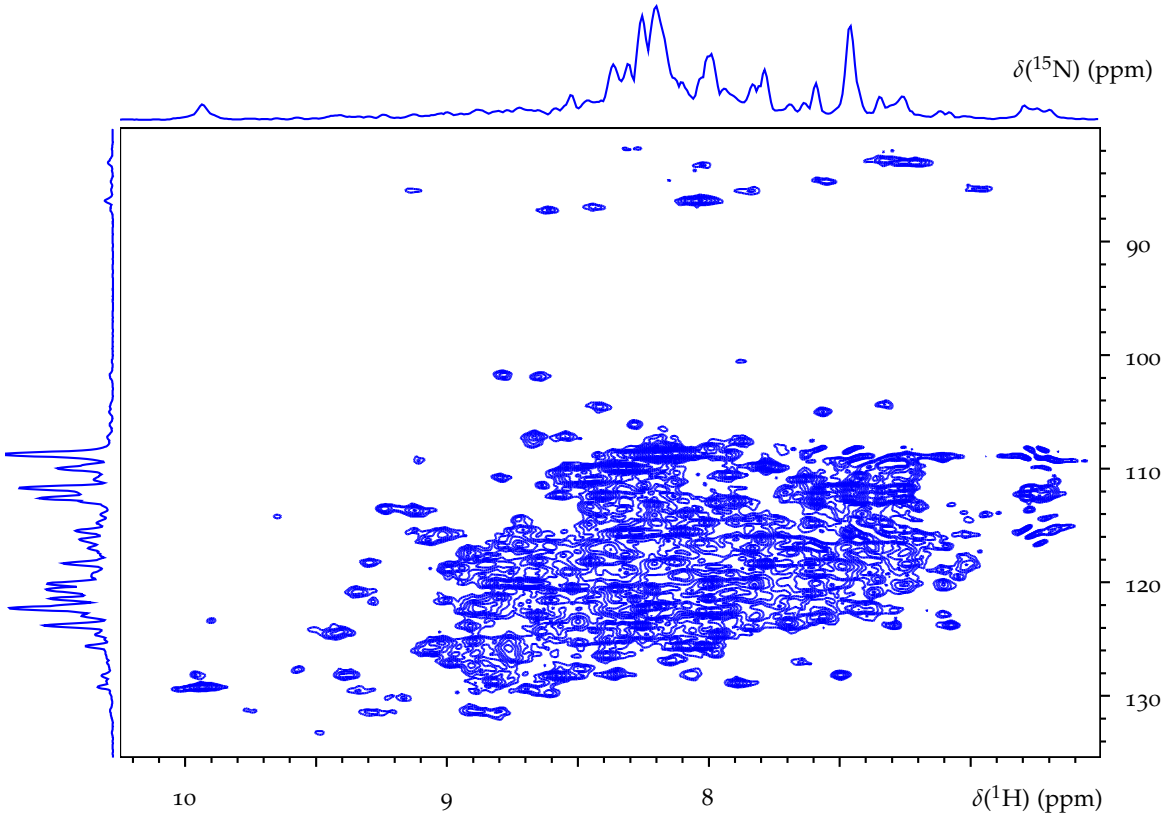


Figure 7.12 – *^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled AP-actin overexpressed in Sf9 cells using tYA$_{10}$-ESF medium. The NMR spectrum was recorded at a 950MHz and 298K.*

ments reduced to their diagonal (equivalent to 2D-NOESY spectra with a null mixing time), ^{15}N decoupled during the non-physical t1 labeling period but not during the t2 acquisition times. ^{15}N isotopic incorporation is obtained by comparison of the three peaks observed for each frequency in the ^{1}H dimension, the central peak corresponds to ^{1}H bound to ^{14}N and the peripheral correspond to ^{1}H bound to ^{15}N. Incorporation was 63% for actin expressed in Sf9 cells and 66% for DIII-HaV and DIII-RVFV expressed in S2 cells.

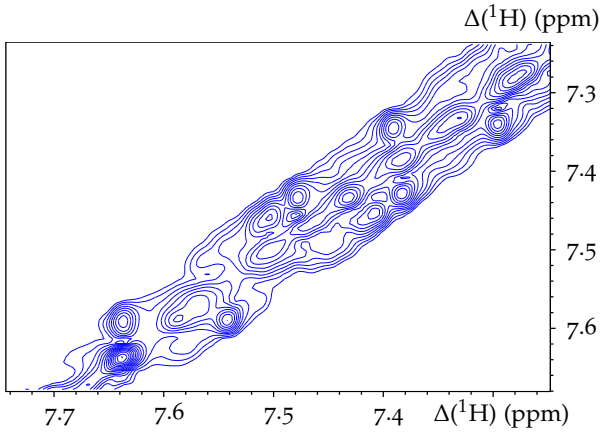


Figure 7.13 – *Correlation spectrum used to evaluate the ^{15}N incorporation ratio in actin.*

Three reasons could explain this rather low incorporation ratio. 1) ESF921-Δaa is not completely amino acid free. 2) The cells still contain unlabeled amino acids when transferred in the labeling medium. 3) Amino acids are synthesized from unlabeled sources during the production step. To test the second hypothesis, S2 cells were transferred in ^{15}N-labeled tYA$_{10}$-ESF medium two days before induction and transferred again in fresh medium just before induction. This additional step enabled to increase the incorporation ratio up to 77% and 80% in the two assay performed but at the cost of a doubling of the labeling medium volume required. Another solution could be to incubate insect cells in ESF921-Δaa to purge the cells from unlabeled amino acid. Preliminary tests showed that it is possible to incubate Sf9 cells in ESF921-Δaa medium for periods up to 24 hours without compromising cell viability however we still have to verify whether this step affects the production yield. Besides, NMR analysis of ESF921-Δaa showed that the medium still contains no negligible amounts of amino acids that will contribute to isotopic dilution.

Structural and dynamic analysis of free WH2 domains

8

This chapter will present the analysis by NMR spectroscopy of the structural and dynamic properties of four free intrinsically disordered WH2 repeats : CH1 and CH2, two chimeras of thymosin $\beta 4$ and Ciboulot of close amino acid sequences and different functions, the WH2 bidomain of the human N-WASP protein and the WH2 bidomain AB of the human Cordon-bleu protein.

8.1 Backbone assignment of free NWASPV2 and CoblAB

The backbones resonances of CH1 and CH2 were assigned by François-Xavier Cantrelle during his PhD.

The (^{1}H-^{15}N)-HSQC spectra of WH2 repeats of N-WASP and cordon-bleu are typical of those of intrinsically disordered proteins with a low spectral dispersion in the amide proton dimension. NMR experiments were carried out at 278K to slow down water amide proton exchange. NWASPV2 and CoblAB resonances were assigned thanks to three dimensional experiments described in the methods section.

The overall backbone assignment of 109-amino acid long CoblAB was completed to 98% for the non-proline $^{1}H^{N}$–^{15}N resonances, with the exception of N-terminal residue, His34 and Ser67 for which signal was too broad due to water amide proton exchange, 99% for the $^{13}C^{\alpha}$, 99% for the $^{13}C^{\beta}$ and 98% for the $^{13}C'$ resonances.

The overall backbone assignment of 67-amino acid long NWASPV2 was completed to 100% for the non-proline $^{1}H^{N}$–^{15}N resonances, with the exception of two N-terminal residues, 95.4% for the $^{13}C^{\alpha}$, 98.4% for the $^{13}C^{\beta}$ and 98.4% for the $^{13}C'$ resonances.

Assigned HSQC spectra of NWASPV2 and CoblAB are shown on figures 8.1 and 8.2 respectively. Chemical shifts tables for these two protein are given in annex.

Temperature shift from 278K to 298K leads to higher linewidth and lower intensities for several peaks especially histidines, serine and asparagines.

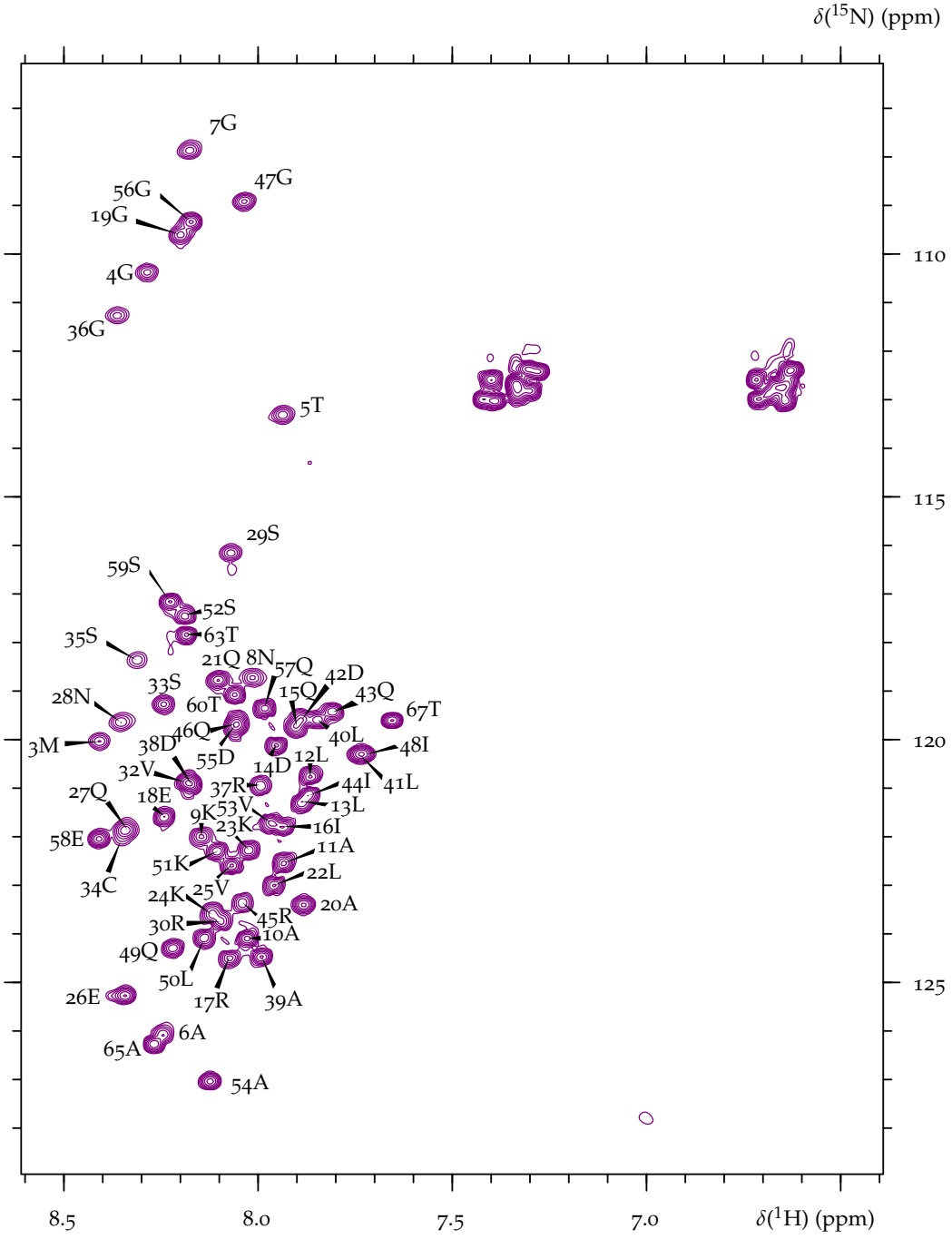


Figure 8.1 – *Assigned (^{1}H-^{15}N)-HSQC spectrum of the WH2 bidomain of human N-WASP in G-buffer (pH=7 & T=278K).*

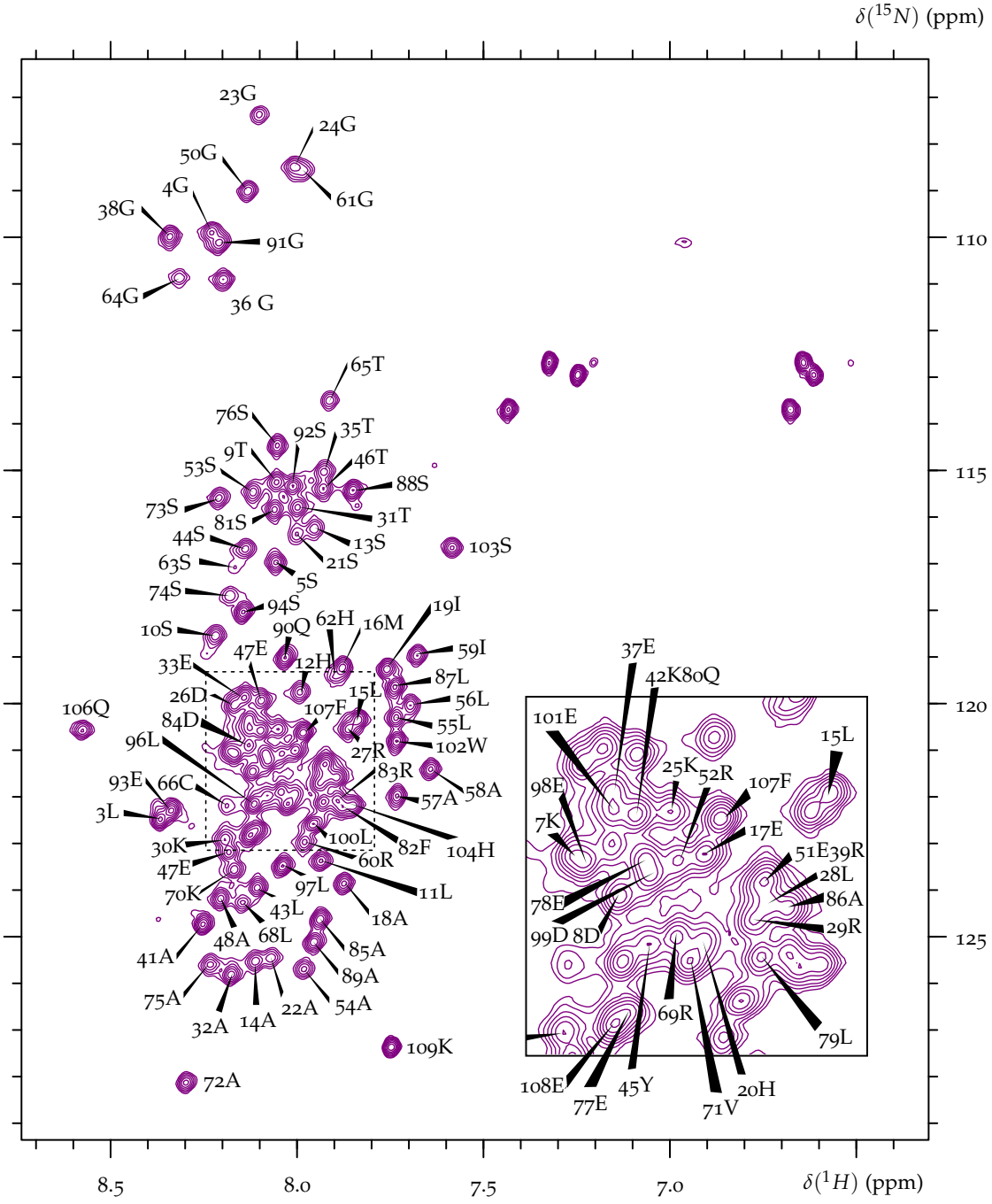


Figure 8.2 – *Assigned (^1H-^{15}N)-HSQC spectrum of the WH2 bidomain AB of human cordon-bleu in G-buffer (pH=7 & T=278K).*

8.2 Residual structure

Intrinsically disordered proteins can be described by ensembles of rapidly inter-converting conformers. Although they lack stable secondary and tertiary structures, a majority of IDPs are not fully random coil and retain residual secondary structure and/or long-range contacts that may be relevant for their functions (Tompa 2005). For instance, in the case of folding upon binding through a conformational selection mechanism, more important residual structure leads to increased binding kinetics and stronger binding affinity (Borcherds et al. 2014). Therefore the description of IDPs ensembles is important to provide insight into function and interaction of theses disordered protein with their partners.

We investigated secondary structure propensities of CH1, CH2, NWASPV2 and CoblAB using chemical shift analysis and residual dipolar coupling.

8.2.1 Chemical shifts analysis

Only carbon chemical shifts were used to evaluate structural content because amide proton and nitrogen chemical shifts are more sensitive to experimental conditions. Carbon chemical shifts were re-referenced using the SSP referencing script (Marsh et al. 2006). SSP scores for the four proteins at 278K, pH=7 and low ionic strength (G-buffer) are represented on figure 8.4. Residual structural propensities evaluated by the δ2d software (Camilloni et al. 2012) are represented of figure 8.4.

The regions which are expected to form α-helices upon actin binding are highlighted in gray. For chimeras, α-helices are identified from solution state NMR data, using chemical shifts index values for $^{1}H^{N}$, ^{15}N and $^{13}C^{\alpha}$ atoms and the TALOS software (Shen et al. 2009). For CoblAB, α-helices are those present on the crystal structure (PBD : 4JHD). For NWASPV2, α-helices boundaries are set using the crystal structure of the NWASPV2:actin (1:1) complex (PDB : 2VCP).

Overall, regions which form α-helices in the actin bound form retain residual helical structure in the free form. Proportion of α-helices in these regions evaluated by chemical shift analysis are summarized in table 8.1. The SSP program predicts higher helical propensities than the δ2d program. SSP scores are globally positive indicating preponderance of helical conformers. Regions with negative SSP scores are centered around Pro10, Pro38 and Pro40 of CH1 and CH2, around Pro31 of NWASPV2, around Pro40 of CoblAB, in the proline-rich region after Thr60 for NWASPV2 and in the Streptag of CoblAB. Prolines are designated by black arrows on figure 8.3.

For chimeras CH1 and CH2, residual secondary structures are very similar except in the linker region. The amphipathic N-terminal helix and linker regions retain more residual α-helical structure than the C-terminal helix region. Helical propensity predicted by the δ2d software is 3% in this region which could be considered as negligible compared to the first amphipathic α-helix.

The residual α-helices of NWASPV2 predicted by the δ2d software do not exactly

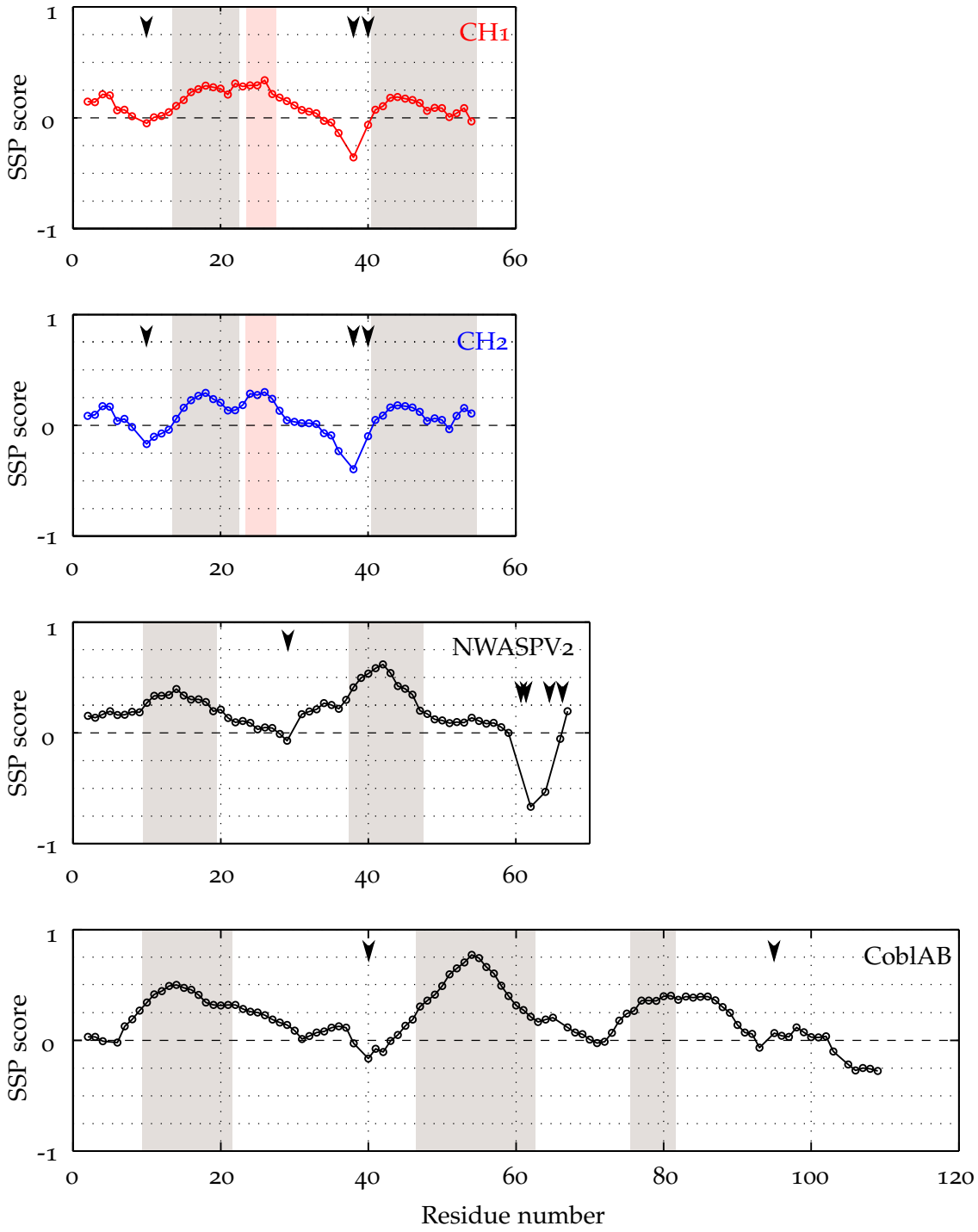


Figure 8.3 – *SSP scores for free CH1, CH2, NWASPV2 and CoblAB. Experiments were carried out at 278K, pH=7 in G-buffer. Positive scores indicate presence of α-helical content and negative scores indicate more extended regions containing β-sheet and polyproline II conformations. Regions which form helices upon actin binding are highlighted in gray. The linker regions of CH1 and CH2 are highlighted in salmon pink. Prolines are designated by black arrows.*

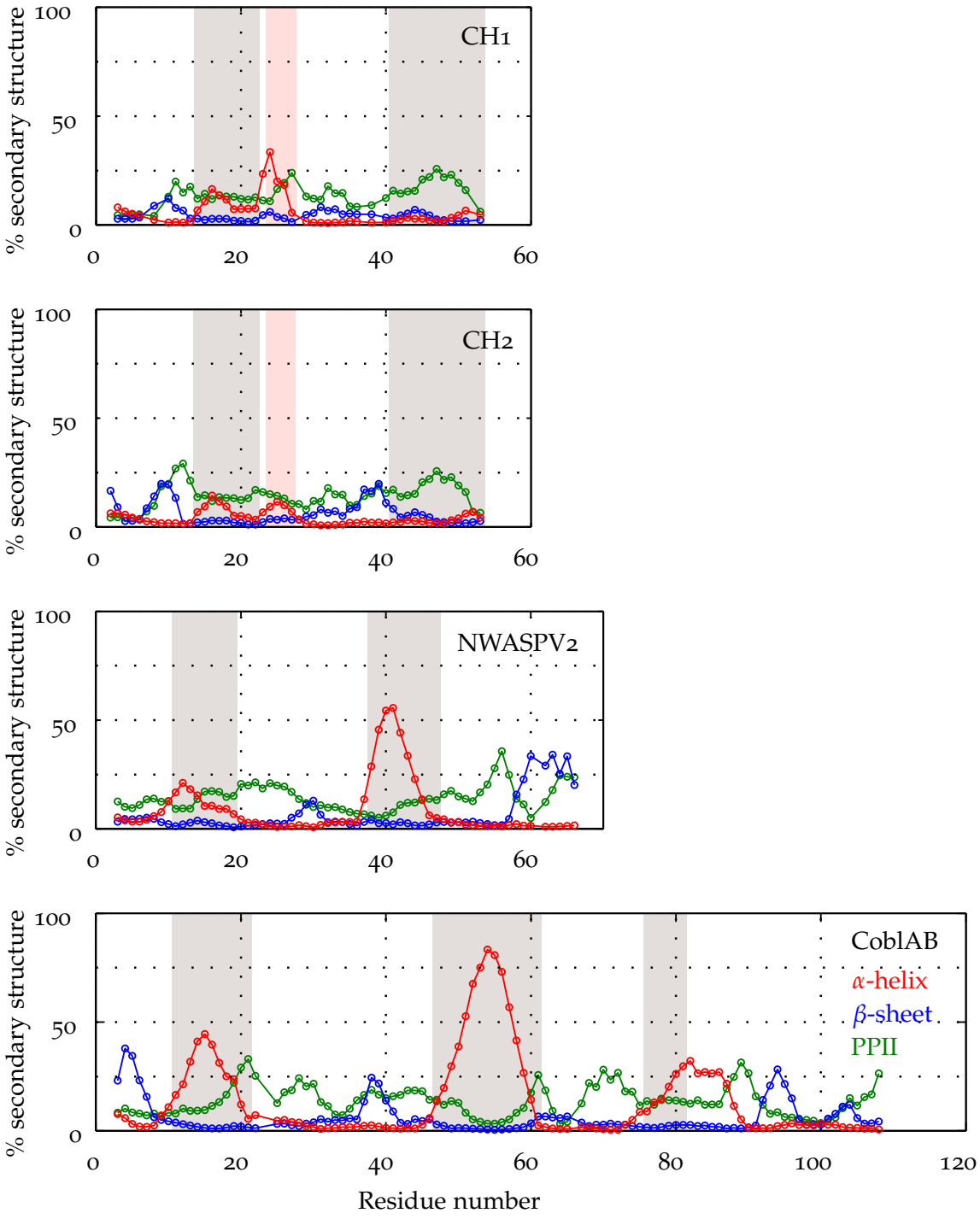


Figure 8.4 – *Secondary structure propensities for CH1, CH2, NWASPV2 and CoblAB using the δ2d software. Propensities to form α-helices, β-sheets and polyproline II secondary structures are described by red, blue and green curves respectively. Experiments were carried out at 278K, pH=7 in G-buffer. Regions which form helices upon actin binding are highlighted in gray. The linker regions of CH1 and CH2 are highlighted in salmon pink.*

have the same limits as in the crystal structure. The residual α-helix of the 1st domain starts at Asn8 (instead of Ala10). The residual α-helix of the 2nd domain goes from Arg37 instead of Asp38 to Arg45 instead of Gly47. Besides, the second amphipathic helix is more populated than the first which could be explained by N-capping by Asp38.

The second residual amphipathic of CoblAB helix only extends up to Arg60. Intriguingly, for CoblAB the C-ter α-helix in the free form is longer than the one observed on the crystal structure and extends from Ser73 to Ala89 (Ser73 to Ser81 in the crystal structure) with an average α-helical propensity of 20% (δ2d) to 31% (SSP).

The CoblAB protein comprises more populated helical conformers. We wanted to know whether it was the result of stabilization of amphipathic helices at low temperature or if these states were also populated at 298K. Carbon chemical shifts at 298K were analyzed with δ2d and similar and equally populated secondary structures were found.

Protein	Secondary structure	SSP	δ2d
CH1	N-ter amphipathic α-helix (14-22)	23%	10%
	linker 3-10 α-helix (24-27)	28%	19%
	C-ter α-helix (41-54)	10%	3%
CH2	N-ter amphipathic α-helix (14-22)	21%	8%
	linker 3-10 α-helix (24-27)	27%	9%
	C-ter α-helix (41-54)	10%	3%
NWASPV2	1st amphipathic α-helix (10-19)	31%	13%
	2nd amphipathic α-helix (38-47)	45%	31%
CoblAB	1st amphipathic α-helix (10-21)	38%	25%
	2nd amphipathic α-helix (47-61)	57%	42%
	C-ter α-helix (73-81)	36%	19%
	(73-89)	30%	20%

Table 8.1 – *Residual secondary structures in free WH2 repeats (278K, pH=7, G-buffer) evaluated by carbon chemical shift analysis in regions which form α-helices in the actin bound form.*

8.2.2 Residual dipolar couplings

Residual secondary structures were also analyzed using residual dipolar coupling data.

$^1D_{HN}$, $^1D_{COCA}$ and $^1D_{COCA}$ residual dipolar couplings were measured for CH1, CH2, CoblAB and NWASPV2 using experiments described in the method section. RDC values depend on the degree of alignment of the sample. To avoid any confusion, all RDC values

mentioned hereafter are for an aligned sample with a deuterium quadripolar splitting of 10Hz, unless stated otherwise.

$^1D_{COCA}$ residual dipolar couplings are very small and exhibit only small variation depending on structural elements. For example, figure 8.5.b. shows expected COCA RDCs for completely disordered CH2 compared to CH2 were the N-terminal amphipathic α-helix is fully formed (deuterium quadripolar splitting of 10Hz). COCA RDCs vary by less than 2Hz between these two extreme conformations. Precision on $^1D_{COCA}$ RDCs is not sufficient therefore they were not included in the analysis for determination of secondary structure propensities.

On the other hand, HN and CH RDCs are subject to more important variations upon structural changes as shown on figures 8.5.a. and c. HN and CH RDCs vary by about 15 and 10Hz between disordered and α-helical conformers.

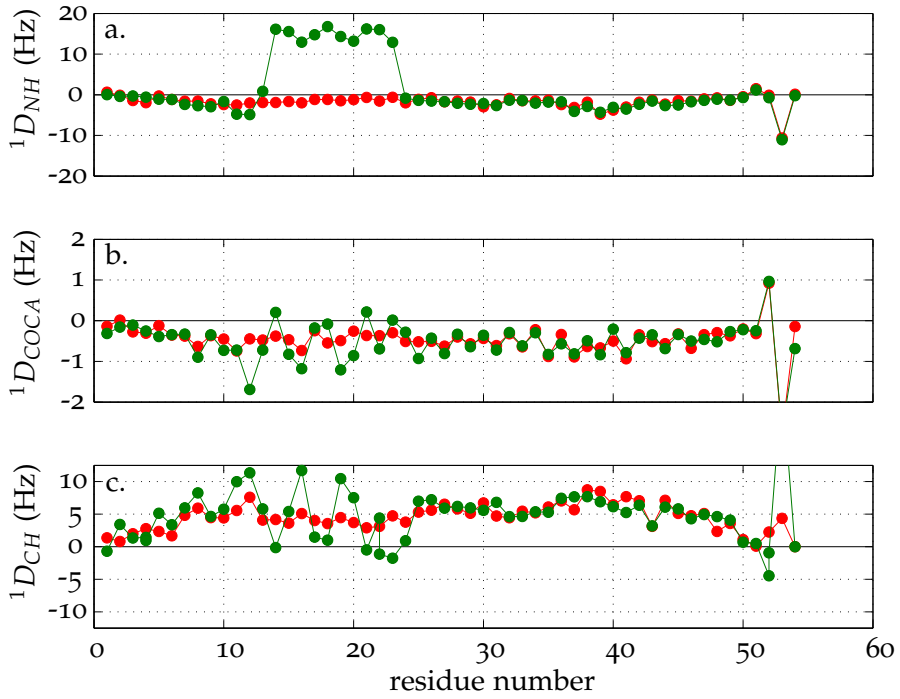


Figure 8.5 – *Expected HN (a.), CACO (b.) and CAHA (c.) RDC values for completely disordered CH2 (in red) compared to CH2 were the N-terminal amphipathic α-helix is fully formed (in green). RDCs were simulated using* flexiblemeccano *for a deuterium quadripolar splitting of 10Hz.*

8.2.2.1 Chimera of Tβ4 and CibD1

Figure 8.6 shows RDCs measured for CH1 and CH2 in PEG/octanol. As for chemical shifts, little differences are observed between the two chimeras, except in the linker region and in the amphipathic helix where HN RDCs are slightly lower for CH2 compared to CH1.

For a random coil protein, lacking residual secondary structures and long range contacts, HN and CH RDCs are expected to display an inverted bell-shaped and a bell-shaped

baseline respectively with local variation linked to amino acid sequence. Deviations from this baseline are observed in two regions. From residues 15 to 28, HN RDCs values are higher, as expected for a region with helical propensity. Around the (Pro38Leu39Pro40) motif, HN RDCs are lower than the baseline as expected for a more extended region.

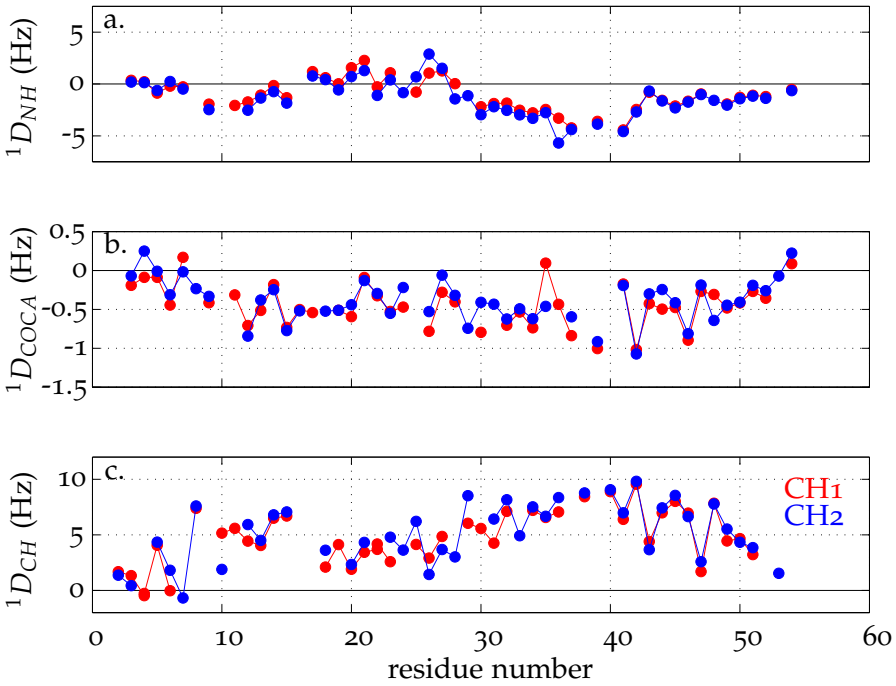


Figure 8.6 – *Experimental HN, CH and COCA RDCs for CH1 and CH2. Estimated uncertainties are around 0.5Hz.*

Conformers ensembles were generated using the Flexible-Meccano software, as described in the experimental section, to provide a more quantitative description of structural propensities using RDC data. A total of 100 000 conformers were generated for each ensemble. RDC values calculated for ensembles generated from 10 000 conformers (default value of the software) were not reproducible.

Experimental RDCs of CH1 and CH2 were first compared with simulated values for a fully disordered protein without any secondary structure propensity (red curves on figure 8.7 and 8.8). Simulated and experimental HN RDC values are in good agreement in the C-terminal region of the protein which forms the C-terminal α-helix in the actin bound form. This is consistent with δ2d prediction of a disordered C-terminus with only 3% of helical conformers. However, experimental HN RDCs for residues 15 to 28 are significantly higher than expected for a fully disordered protein. This is also consistent with higher proportions of helical conformers in the amphipathic helix and linker regions predicted by δ2d.

Experimental RDCs were then compared with simulated values for a conformer ensemble with helical propensities predicted by δ2d (blue curves on figure 8.7 and 8.8) : 10% in the amphipathic N-terminal α-helix (residues 14 to 22) and 19% in the linker (3-10-helix,

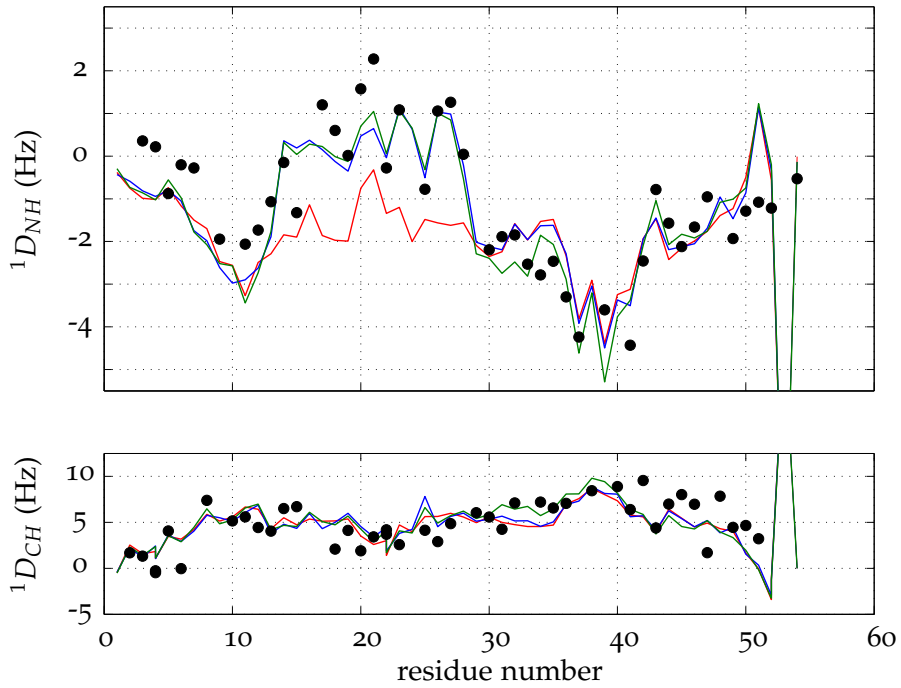


Figure 8.7 – *Experimental HN (top) and CH (bottom) RDCs compared with expected values for fully disordered CH1 (in red), CH1 with helical propensities of 10% in the amphipathic helix and 19% in the linker (in blue) and CH1 with helical propensities of 10% in the amphipathic helix and 19% in the linker and 5% β-sheet propensity in the central region (in green). RDCs were simulated using* flexiblemeccano *for a deuterium quadripolar splitting of 10Hz.*

residues 24 to 27) for CH1 and 8% in the amphipathic N-terminal α-helix (residues 14 to 22) and 9% in the linker (3-10-helix, residues 24 to 27) for CH2.

In the case of CH1, good agreement between simulated and experimental HN RDCs is recovered for residues 22 to 27. However experimental HN RDCs of Lys17, Ser18, Leu20 and Glu21 are higher by 0.5 to 1 Hz than simulated values and experimental HN RDCs of Thr33, Gln34, Glu35 and Lys36 are higher than experimental values. Simulated RDCs for a conformer ensemble with a higher helical propensities of 13% in the amphipathic N-terminal helix (residues 14 to 22) and 19% in the linker region were not in better agreement with experimental data (not shown). The best description of experimental data was obtained for an ensemble of conformers with helical propensities of 10% in amphipathic N-terminal helix (residues 14 to 22) and 19% in the linker (3-10-helix, residues 24 to 27) and a β-sheet propensity of 5% for residues 32 to 38 (green curves on figure 8.7).

In the case of CH2, simulated RDCs for helical populations determined using δ2d are still smaller than experimental RDCs in the amphipathic helix and linker regions (dark blue curves on figure 8.8). The cyan curves on figure 8.8 show simulated RDCs for a conformer ensemble with helical propensities of 13% in the amphipathic N-terminal helix (residues 14 to 22) and 25% in the linker region. These values are in better agreement with experimental data. Finally, a β-sheet propensity of 5% for residues 32 to 38 was finally added to describe smaller HN RDCs of residues 32 to 36 (green curves on figure 8.8).

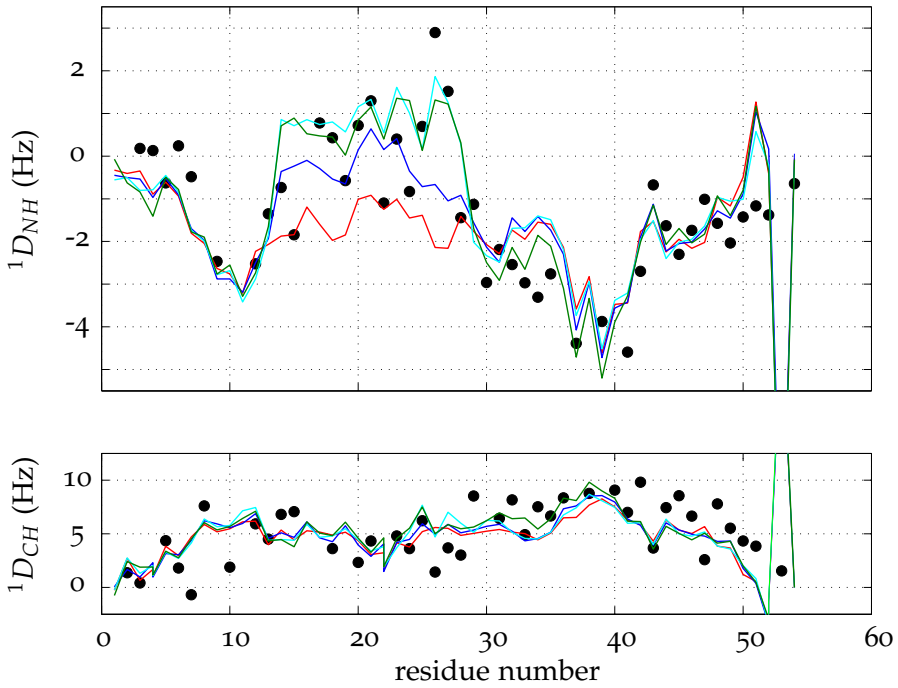


Figure 8.8 – *Experimental HN (top) and CH (bottom) RDCs compared with expected values for fully disordered CH2 (in red), CH2 with helical propensities of 8% in the amphipathic helix and 9% in the linker (in blue), CH2 with helical propensities of 13% in the amphipathic helix and 25% in the linker (in blue) and CH2 with helical propensities of 13% in the amphipathic helix and 25% in the linker and 5% β-sheet propensity in the central region (in green). RDCs were simulated using* flexiblemeccano *for a deuterium quadripolar splitting of 10Hz.*

The *flexiblemeccaco* software does not allow one to launch a series of simulations for different conformational ensembles so all simulations were simulations were started manually which made it difficult to test systematically all residual secondary structure possibilities.

Description of experimental RDCs could probably be further improved by taking into account an equilibrium between several conformations with different secondary structure elements and different population. This approach was especially used for IDPs with high helical content (Jensen et al. 2008, Kragelj et al. 2015). Here the population of helical conformers is about 10-20% compared to 65-75% for the above references. Therefore RDCs variations are also smaller and to avoid overfitting, we limited our description to an equilibrium between an unfolded conformation and one minor conformation with secondary structure elements.

8.2.2.2 WH2 bidomains NWASPV2 and CoblAB

The same approach was applied to the analysis of HN and CH RDCs of WH2 bidomain of human N-WASP. Experimental RDCs are shown in black on figure 8.9. Once again there are significant differences with simulated RDCs for a fully disordered protein lacking secondary structure and long range contacts. The blue curves show simulated RDCs

including the structural propensities predicted by δ2d. The region between the two amphipathic helices has higher HN RDCs than expected for a disordered region. Since no helical propensity is expected for this region, RDCs were simulated including a long range contact between the two amphipathic helices (green curves).

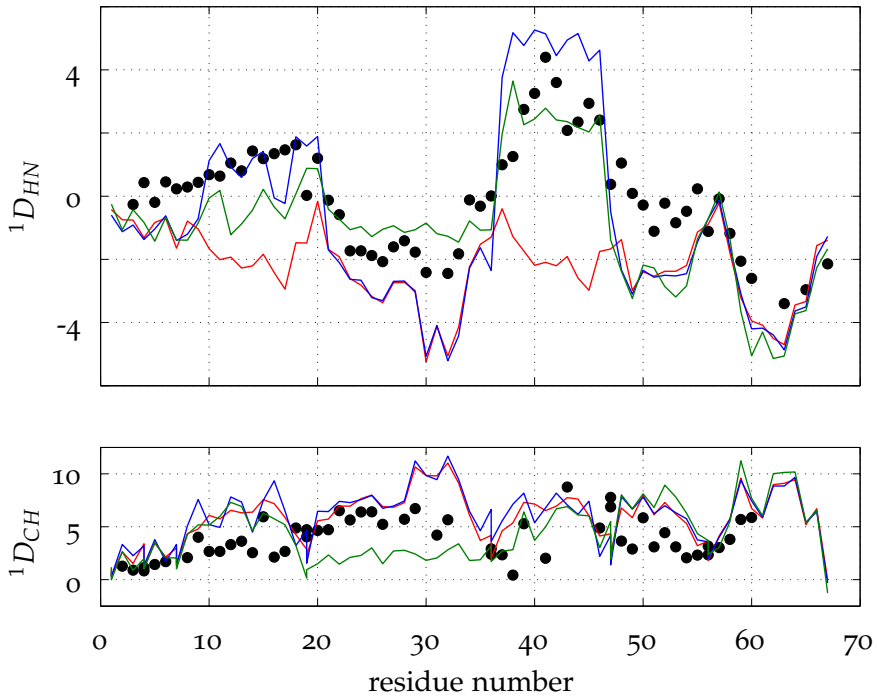


Figure 8.9 – *Experimental HN (top) and CH (bottom) RDCs compared with expected values for fully disordered NWASPV2 (in red), NWASPV2 with helical propensities of 13% in the first amphipathic helix and 31% in the second amphipathic (in blue) and NWASPV2 with helical propensities of 13% in the first amphipathic helix and 31% in the second amphipathic with long range contacts between the two helices (in green). RDCs were simulated using* flexiblemeccano *for a deuterium quadripolar splitting of 10Hz.*

These possible long range contacts can be confirmed and characterized by paramagnetic relaxation enhancement (PREs) experiments. For this experiment, a paramagnetic label, usually a nitroxide group, is covalently attached to a cystein sidechain. The gyromagnetic ratio of the electron spin is 658 times higher than the proton spin so relaxation rates of spatially close protons are dramatically enhanced allowing identification of transient contacts between the region of the paramagnetic probe and the rest of the protein.

CoblAB RDCs were measured using stretched polyacrylamide gels because the protein interacted with the PEG/octanol mixture. The degree of alignment for this sample was less important : HDO quadripolar splitting was between 10 and 12Hz, 2 to 4 times lower than that of samples aligned in PEG/octanol mixtures. Therefore the precision on RDC measurements was 2 to 4 times lower. Besides, the sample suffered partial degradation during re-hydration of the gel and degradation peaks complicated spectra analysis. Reliable HN RDC values were obtained by comparison of the 2D-HSQC-IPAP and 3D-

HNCO-JHNscale experiments. CH RDCs contained too many outliers therefore they were excluded from the data analysis.

Experimental HN RDCs are shown in black on figure 8.10. There are significant differences with simulated RDCs for a fully disordered protein lacking secondary structure and long range contacts. RDCs were simulated including the structural propensities predicted by δ2d. Overall, increase in the simulated RDC values for the helical regions are in agreement with experimental data. However, there are still several differences with experimental data. For example, in the second amphipathic helix, simulated RDCs are in good agreement with experimental data for residues 53 to 58 for which RDC values are between 6 and 10Hz. But experimental RDCs seem to be in agreement with a shorter helix as they decrease from residue 59. Besides residues 47 to 49 form a first plateau with HN RDC values between 1.3 and 2.5Hz which could be consistent with a lower helical population in this region.

Moreover, experimental RDCs between helical regions have negative values between -5 and -10Hz which could be compatible with more extended regions.

Description of experimental RDCs could be further improved by including several conformers in equilibrium. Because of the weak alignment for RDCs measurement and partial degradation of the sample, error on RDCs is more important and we decided keep this simplified description to avoid overfitting.

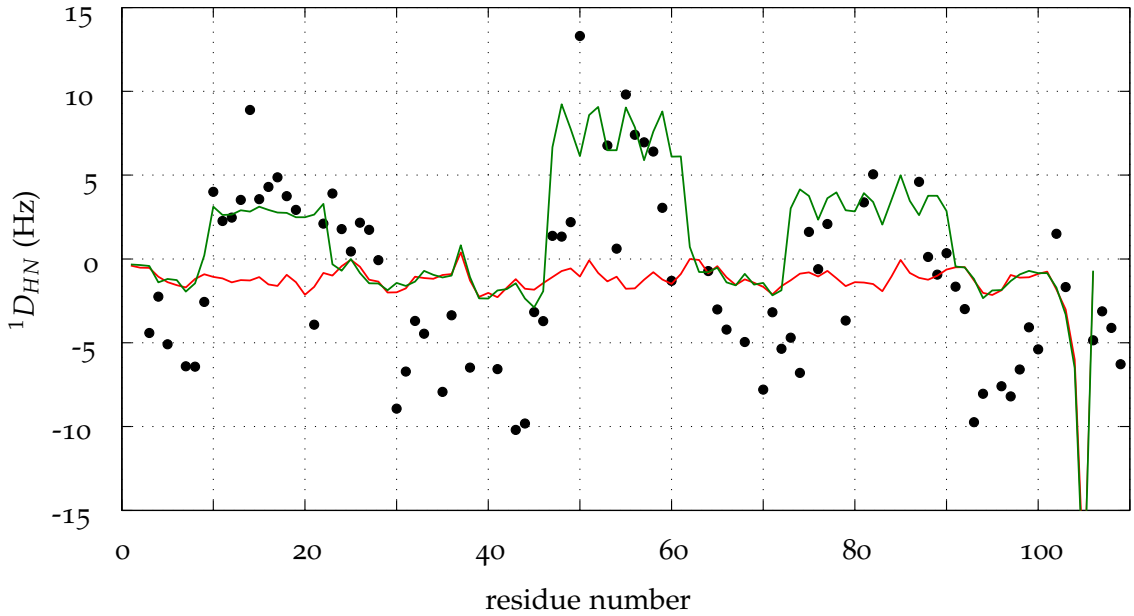


Figure 8.10 – *Experimental HN RDCs (estimated uncertainties are around 1-2Hz) compared with expected values for fully disordered CoblAB (in red) and CoblAB where the two amphipathic α-helices and the C-terminal α-helix are formed with propensities of 25, 42 and 20% respectively (in green). RDCs were simulated using* flexiblemeccano *for a deuterium quadripolar splitting of 10Hz.*

As explained in the experimental section, chemical shifts and RDCs do not average on the same time range. RDCs are especially sensitive to backbone dynamics up to 50μs.

In the case of folded proteins, order parameters S^2_{RDC} can be extracted from RDC data to characterize protein dynamics in the supra τ_C time range (Lakomek et al. 2008). The differential dynamic averaging of chemical shifts and RDCs may explain differences in secondary structure populations determined using these two methods.

8.3 Free WH2 repeats dynamics

8.3.1 Chimera of Tβ4 and CibD1

^{15}N relaxation parameters were measured for free CH1, CH2, NWASPV2 and CoblAB to investigate the backbone motions on the picosecond to nanosecond timescale.

8.3.1.1 Fast backbone dynamics - ^{15}N-relaxation experiment

^{15}N-relaxation experiments were recorded for CH1 and CH2 at 278K, pH=7 in G-buffer at ^{1}H frequencies of 600, 800 and 950MHz, for a detailed analysis of the fast backbone motions in these intrinsically disordered proteins. Figure 8.11 shows experimental ^{15}N-relaxation parameters measured at 950MHz for CH1 and CH2. Profiles are very similar for the two proteins which indicates a similar dynamic behavior in the picosecond to nanosecond time scale. Variations of longitudinal relaxation rates and heteronuclear nOes along the sequence are small except at the extremities of the protein. A few isolated residues (Ala13, Glu43 and Glu46) have higher transverse relaxation rates which might indicate local chemical exchange. Indeed, Ala13 is located at the N-terminus of the residual amphipathic helix. Glu43 and Glu46 have negatively charged side chains and might form transient electrostatic interactions with lysines of the sequence.

The simplest way to describe transverse relaxation for a disordered or denatured protein is by a model where all neighboring residues contributions decay exponentially as the distance from a given residue increases (Schwalbe et al. 1997). The same intrinsic relaxation rate (R_2^{in}) is taken for all residues and the decay of the neighbors contribution is characterized by a persistence length (λ) :

$$R_2(i) = R_2^{in} \sum_{j=1}^{N} exp\left(\frac{\mid i-j \mid}{\lambda}\right)$$

where N is the number of amino acids in the protein.

Adjusted R_2^{in} and λ for CH1 and CH2 were 0.4Hz and 7.1 respectively and calculated transverse relaxation is the black bell-shaped curve on figure 8.11. There are significant differences between this simple model and experimental data, in particular in regions with helical propensities : experimental R_2 are higher by a few hertz in the N-terminal amphipathic helix and linker regions. Additionally, some isolated residues (Ala13, Glu43 and Glu46 designated by black arrowheads on figure 8.11) also display higher R_2 values which might be linked to local chemical exchange.

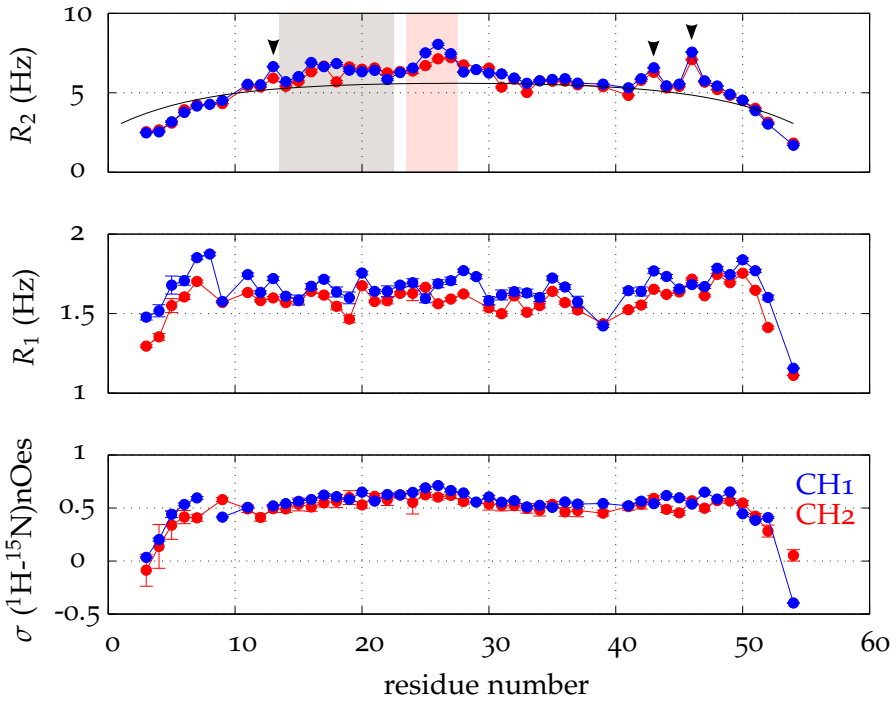


Figure 8.11 – *Comparison of ^{15}N-relaxation parameters for CH1 and CH2 at 950MHz, 278K, pH=7. Uncertainties are generally below 5% of R_1 and R_2 values and below 15% of nOe values.*

The contributions of slow chemical exchange processes in the micro to millisecond timescale were evaluated using the method of Phan et al. (1996) (detailed in the experimental section). However, the exchange contribution did not comply with a fast exchange hypothesis for most residues indicating a more complex exchange contribution.

An intrinsically disordered protein is expected to encounter a various range of conformational transitions associated with conformers of various correlation times. In agreement with this statement, ^{15}N-relaxation data could not be described using a local Lipari and Szabo approach associated with only one correlation time τ_C per residue to account for all conformational transitions in the nanosecond time range and an order parameter S^2 for picosecond libration motions (LS1). Introduction of a second correlation time to account for local fast motion improved the fit but the description of experimental data was still not achieved (LS2). Addition of an exchange contribution $R_{ex} = A_{ex} \times B_0^2$ (LS3 and LS4) improved the fit but failed to describe transverse relaxation (see figure 8.13) indicating that we are not in a fast exchange regime.

The best result was obtained using the LS5 model-free approach where the adjusted parameters are the correlation time τ_C, the order parameter S^2 and α and A_{ex} which define the exchange parameter $R_{ex} = A_{ex} B_0^{\alpha}$ (Millet et al. 2000). The obtained parameters for CH1 and CH2 are shown on figure 8.14.

The average local correlation times for CH1 and CH2 are 1.1ns which is less than expected for a globular 6kDa protein. This is consistent with the idea that this correlation time does not represent overall tumbling of the protein but rather segmental diffusion. The average order parameter value, equal to 0.7, corresponds to fast libration motions

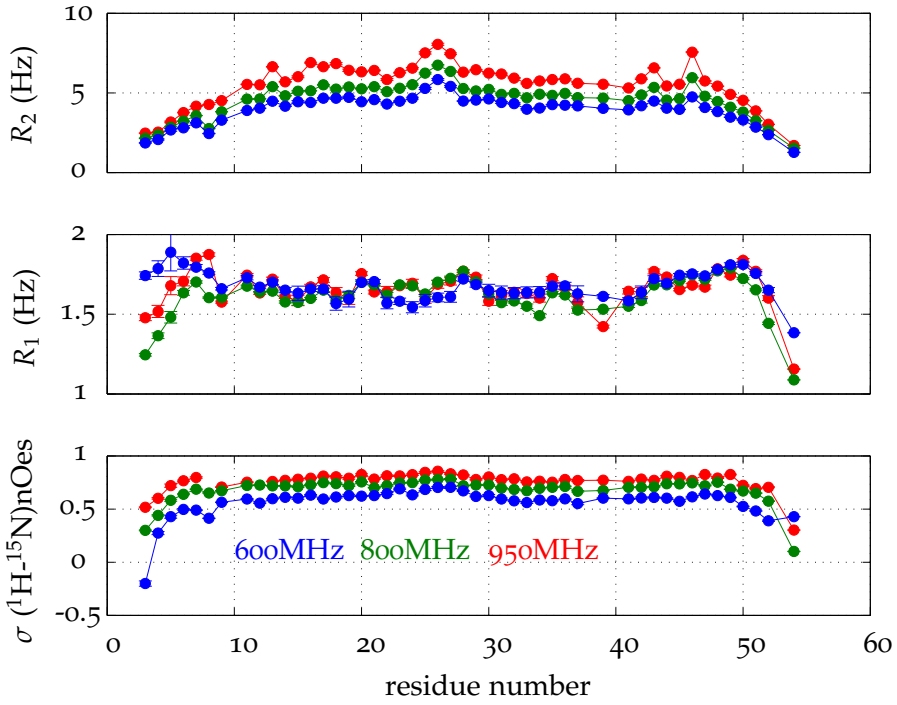


Figure 8.12 – *Longitudinal ^{15}N-relaxation R_1, transverse ^{15}N-relaxation R_2 and heteronuclear ($^1H \rightarrow ^{15}N$) nOes of CH2 at 278K and pH=7, recorded at three different fields.*

in a cone with a half-angle of 25°. The average α value is 1 which is characteristic of an intermediate exchange regime ($k_{ex} \simeq \Delta\omega$). Indeed, the approach of (Phan et al. 1996) could not work to extract R_{ex}. The exchange contribution represents 60 to 70% of the total transverse relaxation R_2.

The use of a lorentzian distribution of correlation time previously allowed the description the heterogeneous dynamics of partially unfolded proteins (Ochsenbein et al. 2002). Description of the relaxation rate was slightly better using this method. χ^2_{red} values were lower for 41 and 28 out of 46 residues for CH1 and CH2 respectively. Addition of an order parameter S^2 did not improve the fit and even lead to higher χ^2_{red} values. However, adjusted correlation time distributions are very broad. The average value for center of the distribution τ_0 is 0.33ns for the two proteins when the average value for the width of the distribution Δ is 0.9ns. This kind of model which implies that the conformational transitions are fluctuations around a predominant conformation is therefore inadapted for these intrinsically disordered domains for which the ensemble of sampled conformers is much more diverse.

These two last approaches aim to associate a correlation time or a distribution of correlation time to the fast dynamics of the protein backbone without describing the nature on the conformational transition.

Another approach is to fit the experimental data to the theoretical time correlation function associated to a specific conformational transition. Previous studies on our peptides showed that they are mainly disordered with little residual secondary structure.

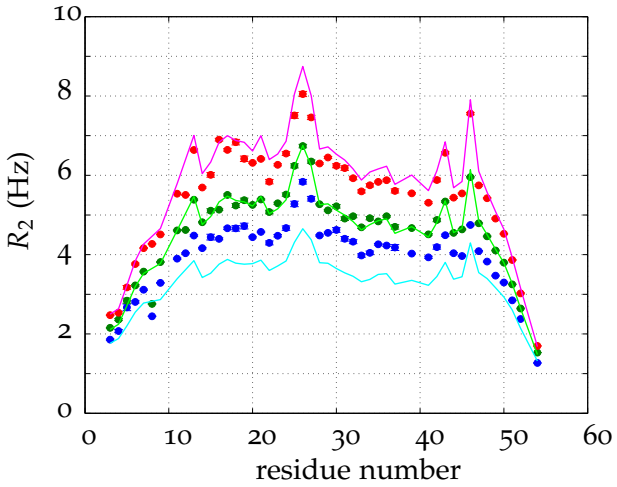


Figure 8.13 – *Comparison of experimental (dots) and calculated (straight lines) R_2 values of CH2 at 600MHz (blue), 800MHz (green) and 950MHz (red). Calculated R_2 were obtained after adjustment of τ_C, S^2 and A_{ex} using the LS3 model.*

Therefore we turned to models first designed for polymer relaxation analysis. Among them, the bimodal time-correlation function developed by Dejean, Lauprêtre, and Monnerie (DeJean de la Batie et al. 1988a;b) (DLM) offers an interesting description of the segmental dynamics of polymers. This model derives from the Hall–Weber–Helfand model (Helfand 1971, Hall and Helfand 1982) in which the backbone dynamics are described by a predominant "crankshaft-like" transition characterized by a correlation time τ_1 and its diffusion process along the backbone characterized by a damping time τ_2. Dejean, Lauprêtre, and Monnerie extended this model to take into account uncorrelated fast HN bond libration in the picosecond timescale by adding an order parameter A_{HH} analog to the order parameter S^2 in the model-free approach. The corresponding expression of the orientation auto-correlation function and spectral density function are given in the experimental section.

This DLM approach was mainly applied to the description of polymer relaxation but was also used to fit C13 relaxation parameters of homo and hetero polysaccharides (Dais et al. 2005). Besides, oscillation of the plane of the peptide group characterized by a "crankshaft-like" motion were already observed in molecular dynamic studies on unfolded peptides (Fitzgerald et al. 2007) although the DLM approach was never used to describe NMR relaxation parameters of IDPs.

Of all our attempts to describe experimental results, this DLM model was the one that best fitted our experimental data. 91% of back calculated longitudinal and transverse relaxation parameters were in agreement with experimental values in a range of 10% for both chimeras. The adjusted DLM parameters are very similar for the two chimeras (see figure 8.15). The correlation times τ_1 and τ_2 follow a bell-shaped curve along the sequence. The mean value for the order parameter A was unusually low : 0.34 for CH1 and 0.33 for CH2 corresponding to a libration demi-cone angle of 47 degrees. Again, this is compatible with the hypothesis of a very flexible and dynamic chain.

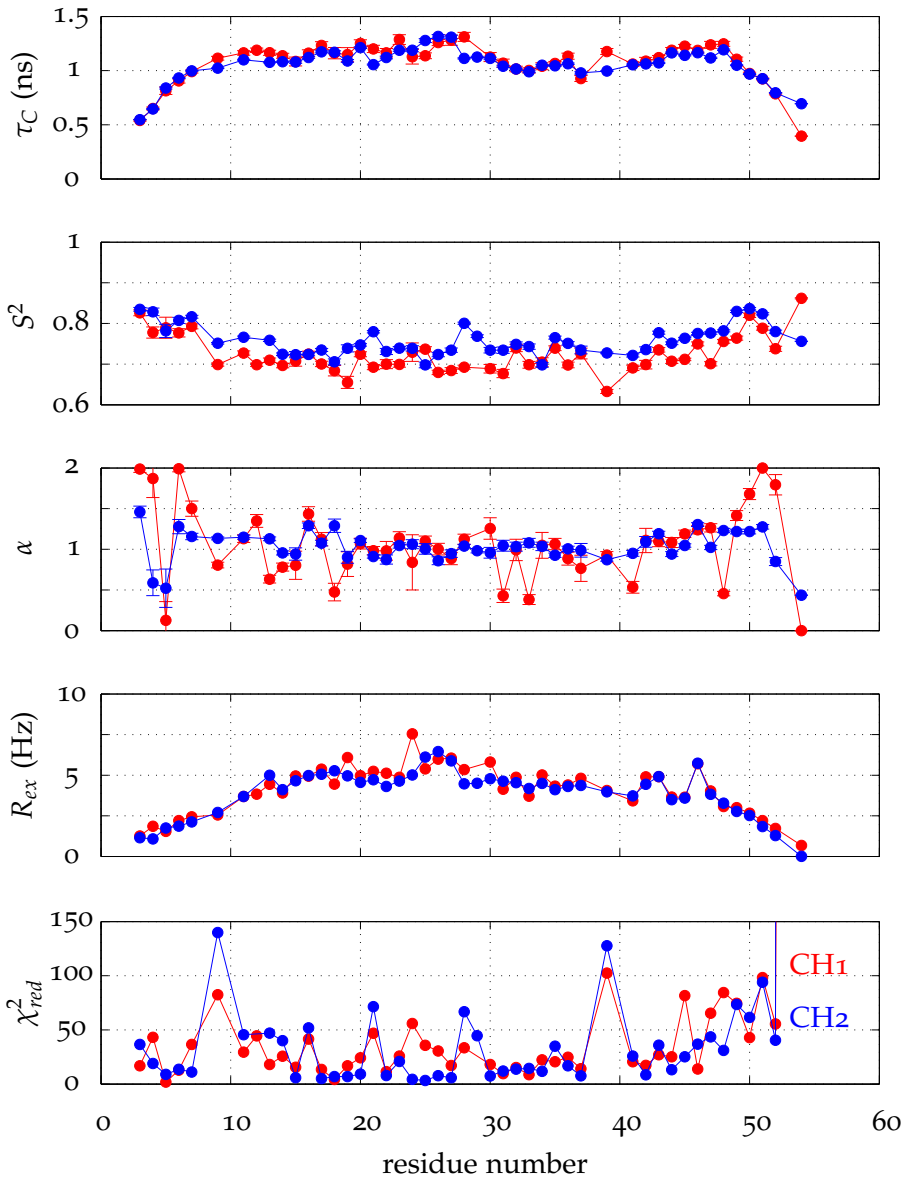


Figure 8.14 – *Dynamical parameters obtained by applying model-free (LS5) analysis to relaxation rates of CH1 and CH2. τ_C and S^2 are local correlation times and order parameters for fast backbone motions. α and A_{ex} quantify slower intermediate conformational exchange R_{ex}. In the intermediate exchange regime, $k_{ex} \simeq \Delta\omega$ so the characteristic times of conformational transitions may have values ranging from the millisecond to the submicrosecond depending on the chemical shift differences between exchanging conformers.*

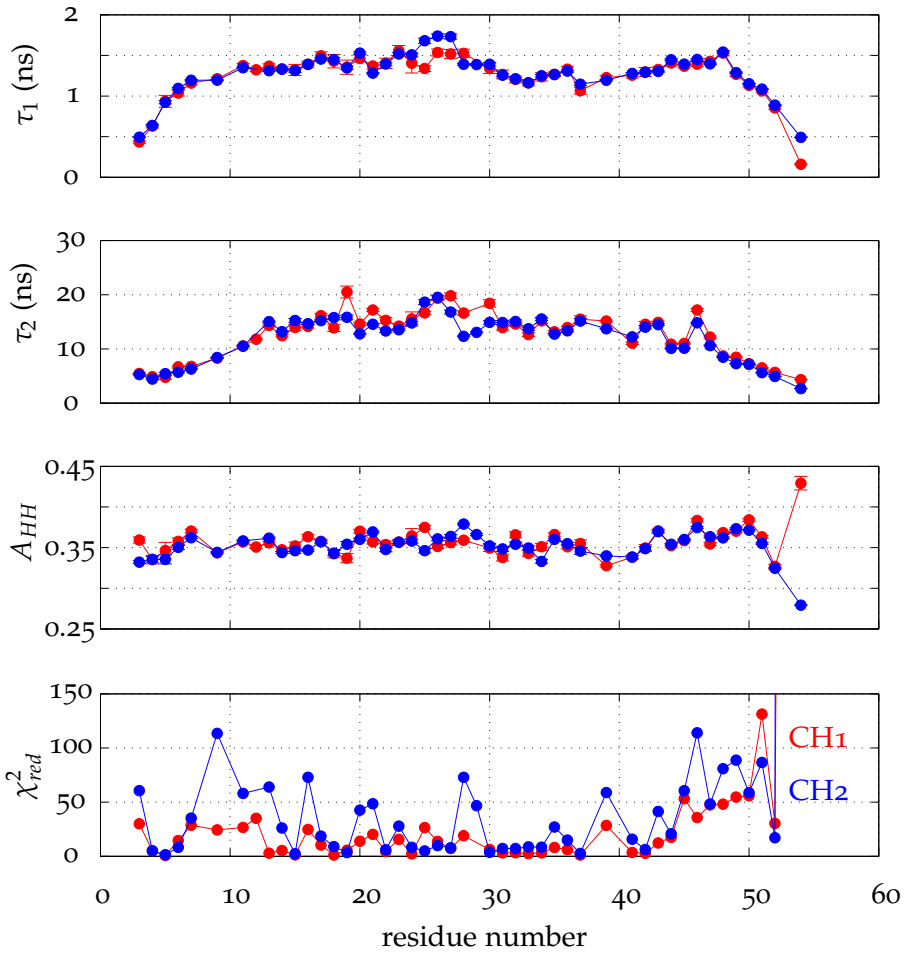


Figure 8.15 – *Dynamical parameters obtained by applying DLM analysis to relaxation rates of CH1 and CH2.* τ_1 *is the characteristic time of the "crankshaft-like" transition,* τ_2 *is the damping time of the diffusion process along the peptide chain and* A_{HH} *is the order parameter linked to the amplitude of fast libration motions.*

In the linker region, τ_1 is higher and τ_2 is lower for CH2. This could be due to the presence of the 3.10 helix turn stabilized by a salt bridge between Asp24 and Lys27, leading to a more rigid environment in which the crankshaft motions are slower and damping is faster. Although this 3-10 helix turn is also present in CH1, it is not stabilized by electrostatic interactions and might be more dynamic.

In conclusion, two models were able to describe the experimental relaxation rates of CH1 and CH2 : a model-free approach including a contribution for slower conformational exchange (LS5) and a model derived from polymer analysis (DLM). These two models offer very different descriptions of the backbone dynamics : on the one hand (LS5), fast libration motions on the picosecond timescale ($S^2 \simeq 0.7$), segmental diffusion with a local correlation time (τ_C) around 1ns and an important contribution of slower intermediate conformational exchange. On the other hand (DLM), there is a more important contribution of fast libration motions on the picosecond timescale and the rest of the backbone

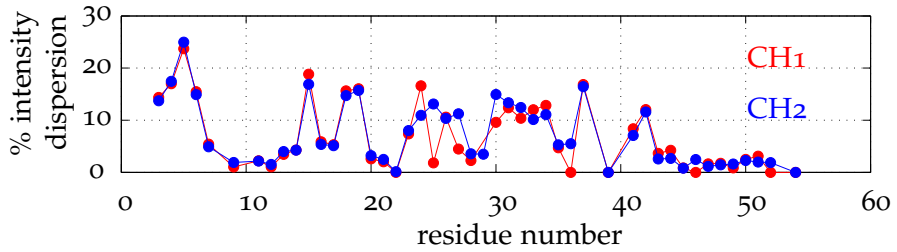


Figure 8.16 – *Relaxation dispersion over the sequence for CH1 and CH2.*

dynamics can be described by a crankshaft-like transition characterized by a correlation time around 1.5ns, with diffusion on the protein backbone characterized by a damping time around 10-20ns and no slower conformational exchange.

8.3.1.2 Slower backbone dynamics and conformational exchange

^{15}N-relaxation dispersion experiments were recorded for CH1 and CH2 with a relaxation delay of 100ms, at 950MHz, 278K and pH=7. Significant intensity dispersion (more than 5%) was observed for 19 and 24 residues of CH1 and CH2 respectively (see figure 8.16). These residues are distributed along the whole sequence.

Relaxation dispersion is due to modification of the electronic environment around the spin in the 100μs to millisecond time scale. It can be due to conformational exchange processes but chemical exchange between water and amide protons may also contribute. To avoid unwanted chemical exchange contributions, we measured relaxation dispersion profiles at 4°C to slow down water/amide proton exchange and used an optimized pulse sequence (Hansen et al. 2008). As a control, we also measured water/amide proton exchange rates using CLEANEX experiments.

Water/amide proton exchange rates measured by CLEANEX experiments are highly correlated to the percentage of relaxation dispersion has shown of figure 8.17. Therefore measured relaxation dispersion seems to be mainly due to proton amide exchange and not conformational exchange. This contribution is lower than expected for a non-optimized relaxation dispersion pulse sequence indicating that it is attenuated by the use ^{1}H continuous-wave decoupling during the ^{15}N CPMG pulse train but not eliminated, probably because of the long relaxation delay of 100ms in our experiments.

Experimental water/amide proton exchange rates were compared with expected values for a fully disordered protein (Bai et al. 1993, Connely et al. 1993). Protection factors which are predicted random coil exchange rate divided by experimental exchange rates are shown on figure 8.18. Most protection factors are around of below 1 even in regions containing residual secondary structure. This shows that residual structures are in exchange with random coil conformations with exchange rates lower than the millisecond. Surprisingly, residues Gln47 to Ala51 have higher protection factors between 1.5 and 2.5 for CH1 and CH2. This region was shown to be completely disordered but transient contacts within this region may partially protect amide protons from exchange with water.

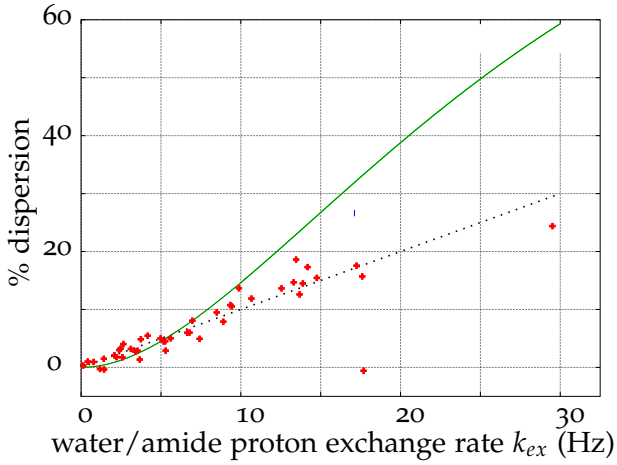


Figure 8.17 – *Relaxation dispersion as a function of water/amide proton exchange rates for CH1 (in red). The green curves shows the expected contribution of water/amide proton exchange to relaxation dispersion for a non-optimized relaxation dispersion pulse sequence (described in section 6.7).*

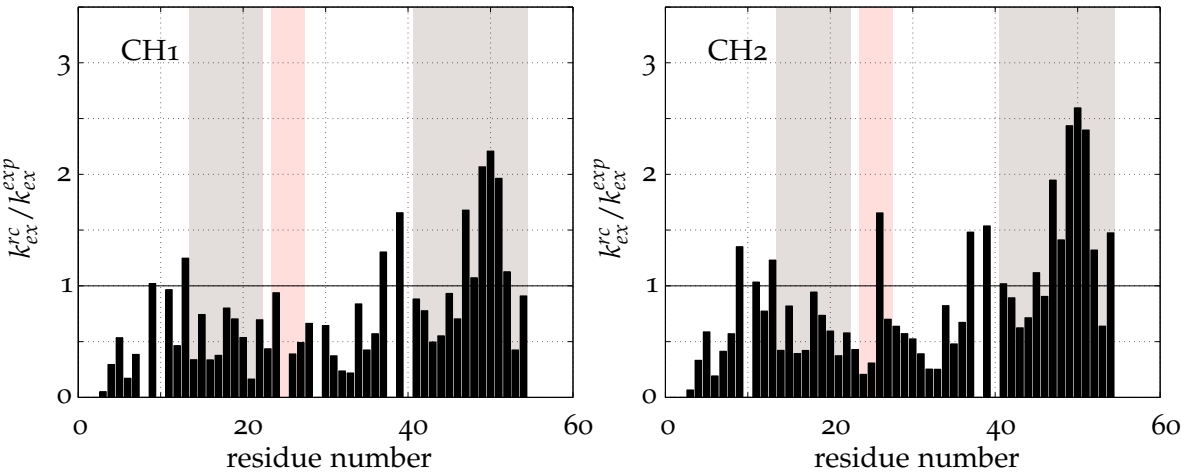


Figure 8.18 – *Protection factors of CH1 and CH2 at 278K at pH=7.*

8.3.2 WH2 bidomains NWASPV2 and CoblAB

8.3.2.1 Fast backbone dynamics - ^{15}N-relaxation experiment

For NWASPV2 and CoblAB, ^{15}N-relaxation parameters were only measured at one magnetic field so analysis of backbone fast dynamics is more qualitative.

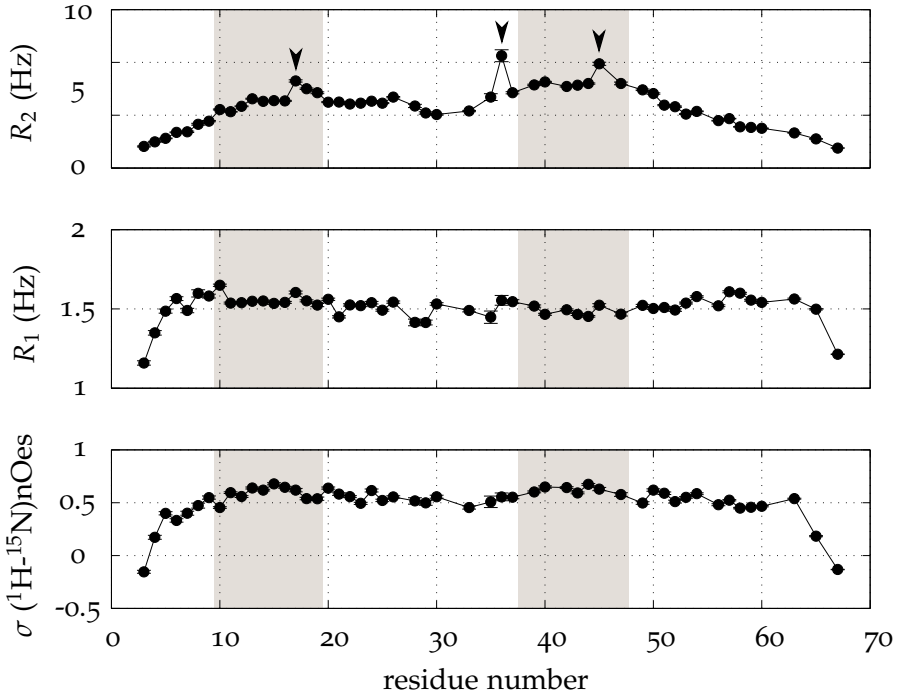


Figure 8.19 – *Longitudinal ^{15}N-relaxation R_1, transverse ^{15}N-relaxation R_2 and heteronuclear ($^1H\rightarrow^{15}N$) nOes of NWASPV2 at 950MHz, 278K and pH=7. Uncertainties are generally below 5% of R_1, R_2 and nOe values The two residual α-helices are shown in gray and residues Arg17, Gly36 and Arg45 which have higher R_2 rates are designated by black arrowheads.*

Figure 8.19 shows the experimental ^{15}N-relaxation parameters measured at 950MHz for NWASPV2. As for CH1 and CH2, the longitudinal relaxation and heteronuclear nOe profiles show little variations. The two regions containing residual α-helices have higher transverse relaxation rates. A plateau with an average R_2 rate of 6.3Hz extends from Leu13 to Glu26 including the residual amphipathic helix of the first domain and the LKKV motif. A second plateau for which boundaries are less clear extends from Ser35 to Leu50, including the residual amphipathic helix of the second domain but only the leucine of the LKSV motif. Additionally, three residues have clearly higher transverse relaxation rates : Arg17, Gly36 and Arg45. Interestingly, Arg17 and Arg45 are conserved residues in WH2 repeats and are also found in human WASP, WAVE, WASH, MIM proteins. This arginine is also present in WH2 repeats of human Spire (replaced by a lysine in the fourth domain) and in the second and third domains of human Cordon-bleu. This increase in transverse relaxation may be linked to conformational exchange in the C-terminus of residual helices.

There is no significant effect of residual structure on longitudinal relaxation rates but heteronuclear nOes are slightly higher in residual helices regions which is consistent with

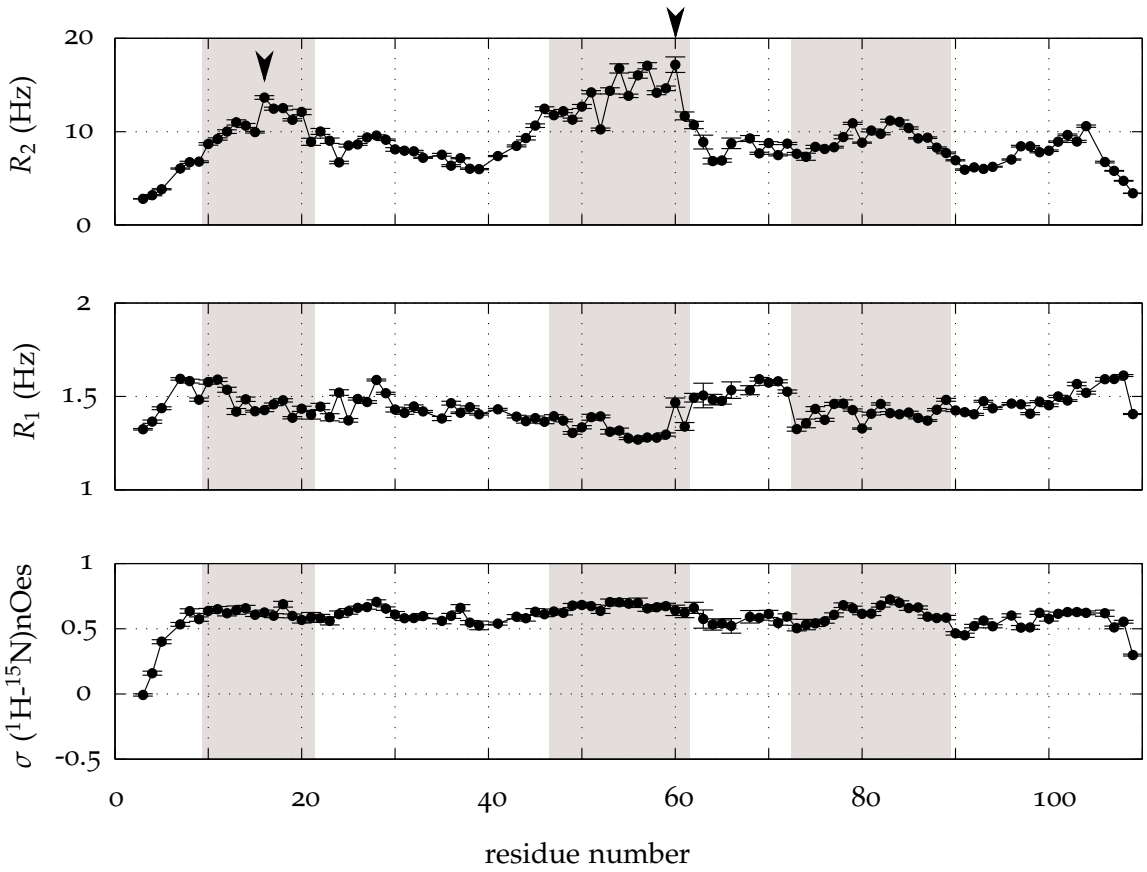


Figure 8.20 – *Longitudinal ^{15}N-relaxation R_1, transverse ^{15}N-relaxation R_2 and heteronuclear (^{1}H→^{15}N) nOes of CoblAB at 950MHz, 278K and pH=7. Residual α-helices are shown in gray.*

reduced flexibility. Average heteronuclear nOe values are 0.59±0.07 and 0.62±0.03 in the two amphipathic helices against 0.51±0.07 for the rest of the protein.

Figure 8.20 shows the experimental ^{15}N-relaxation parameters measured at 950MHz for CoblAB. These relaxation profiles show more important variations, in agreement with higher helical population restraining fast dynamics. As for NWASPV2, the three regions containing residual α-helices have higher transverse relaxation rates. In the case of CoblAB, they also have lower longitudinal relaxation rates and higher heteronuclear nOes especially for the second residual amphiphatic helix. Average heteronuclear nOe values are 0.62±0.03, 0.65±0.04 and 0.61±0.06 in the three helices against 0.57±0.05 for the rest of the protein. This is consistent with reduced flexibility in these regions. Additionally, 6 residues, Gly24 to Arg29 including the leucine and arginine of the LRKT motif of the first WH2 repeat, also have higher heteronuclear nOes. Residues in the second half of the first amphipathic helix (from Met16 designated by a black arrowhead to His20) have higher transverse relaxation rates. Conserved Arg60 in the second amphipathic helix is designated by a black arrowhead.

8.3.2.2 Slow backbone dynamics and conformational exchange

^{15}N-relaxation experiments were recorded for NWASPV2 with relaxation delays of 50 and 100ms, at 950MHz, 278K and pH=7.

Important dispersion is observed throughout the sequence and the distribution of peaks displaying significant dispersion is different than that of chimeras (see figure 8.21 left). Intensity dispersion was compared with experimental water/amide proton exchange rates measured by CLEANEX experiments on the same sample. For most residues, intensity dispersion is more important than that expected for water/amide proton chemical exchange at rates determined by CLEANEX experiments. Relaxation dispersion is in this case linked to conformational exchange.

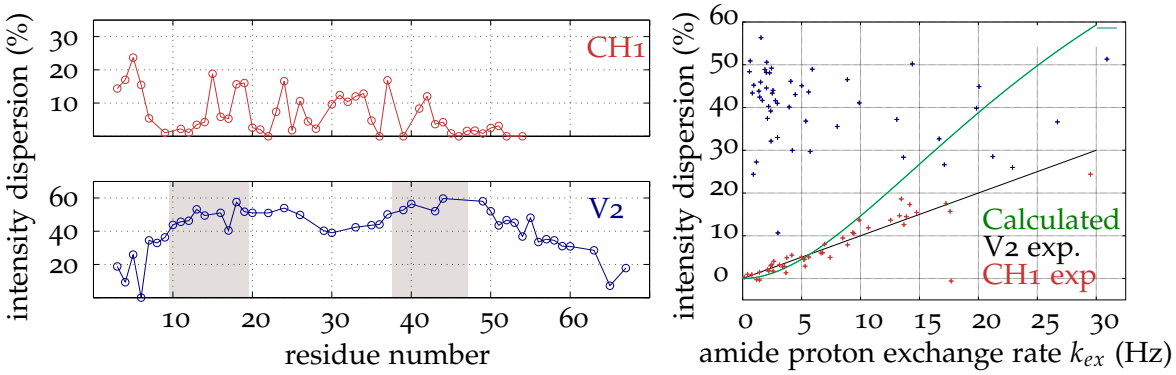


Figure 8.21 – *Left : dispersed intensities in ^{15}N-relaxation dispersion experiments performed on free CH1 (top, in red) and V2 (bottom, in blue). The regions of the amphipathic helices are highlighted in grey. Right : relaxation dispersion as a function of water/amide proton exchange rates for CH1 (in red) and V2 (in blue). The green curve shows the expected contribution of water/amide proton exchange to relaxation dispersion for a non-optimized relaxation dispersion pulse sequence.*

Relaxation dispersion profiles for two relaxation delays were adjusted using a global two-state exchange model. Residues of the first and second amphipathic helix were adjusted separately. The obtained exchange parameters were p_E=9.6%, k_{ex}=28Hz (t_{ex}=35ms) and chemical shifts differences $\Delta\omega \simeq$ 0.5-0.6ppm for the amphipathic helix of the first domain and p_E=10%, k_{ex}=33Hz (t_{ex}=30ms) and chemical shifts differences $\Delta\omega \simeq$ 0.5-0.6ppm for the amphipathic helix of the second domain.

Thus, similar exchange parameters were obtained for the two domains, in agreement with a transition involving the whole fragment. Besides, relaxation dispersion is not restricted to the helix region. Thus it in unlikely, that this measured conformational exchange corresponds to folding/unfolding processes in transiently helical regions but may rather be consistent with exchange between a conformation where the two helices are in contact and a more extended global conformation.

The adjusted k_{ex} is particularly low for relaxation dispersion experiments which are optimal for exchanges rates between 200 and 2000Hz. To confirm this slow-exchange, it would be interesting to measure relaxation dispersion at a lower magnetic field. Indeed, in the slow exchange regime the exchange contribution to transverse relaxation R_{ex} is

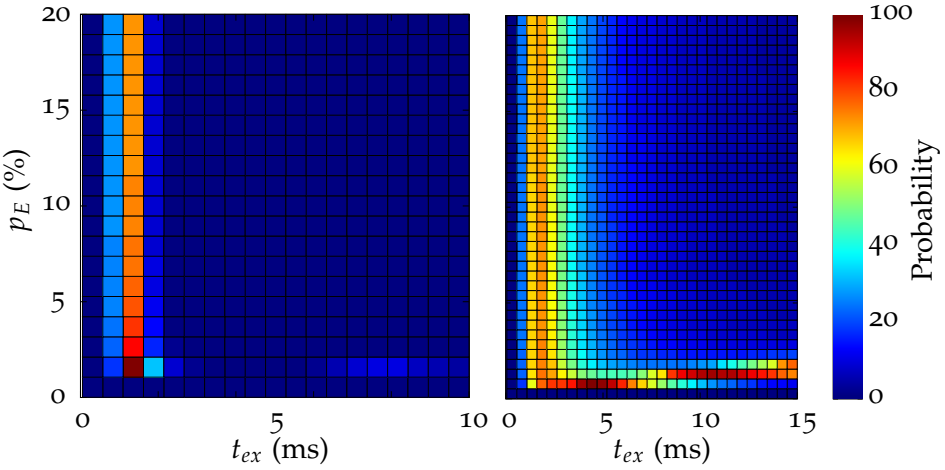


Figure 8.22 – *Adjusted minor state population and exchange rate for residues showing relaxation dispersion in the first (left) and second (right) amphipathic helices of CoblAB.*

independent of the magnetic field B_0. We were not able to record relaxation dispersion experiments at 600MHz because of a pulse sequence incompatibility between the 950 and 800MHz spectrometers on the one hand and the 600MHz spectrometer on the other hand. Pulse, power and delay lists were not incremented properly.

^{15}N-relaxation experiments were recorded for CoblAB with relaxation delays of 40 and 80ms, at 950MHz, 278K and pH=7.

Intensity dispersion was compared with experimental water/amide proton exchange rates measured by CLEANEX experiment on the same sample. Overall, intensity dispersion correlates with that expected for water/amide proton chemical exchange at rates determined by CLEANEX experiments but are higher than expected. Some residues show particularly high intensity dispersion compared to their low water/amide proton exchange (surrounded by the red circle on figure 8.23 left). These residues correspond to the C-terminal regions of amphipathic α-helices : Met16 to His20 and Leu55, Ala57 and Arg60. Relaxation dispersion profiles for two relaxation delays were adjusted using a global two-state exchange model. Residues of the first and second amphipathic helix were adjusted separately : Ser13, Met16, Glu17, Ala18, Ile19 and His20 in the first amphipathic helix and Ser53, Leu56, Ala57, Ala58, Ile59 and Arg60 fo the second amphipathic helix. The obtained exchange parameters were p_E=1-2%, k_{ex}=500-1000Hz (t_{ex}=1-2ms) for the amphipathic helix of the first domain and p_E=1-2%, k_{ex}=100-200Hz (t_{ex}=5-10ms) for the amphipathic helix of the second domain.

Experimental water/amide proton exchange rates were compared with expected values for a fully disordered protein (Bai et al. 1993, Connely et al. 1993). Protection factors for NWASPV2 are shown on figure 8.24. Overall, they are lower than those of CH1 and CH2, indicating that amide protons of NWASPV2 are less accessible. This is consistent with transient contacts suggested by RDCs and relaxation dispersion. Surprisingly, protection factors are lower in residual α-helices and higher in other regions of the protein.

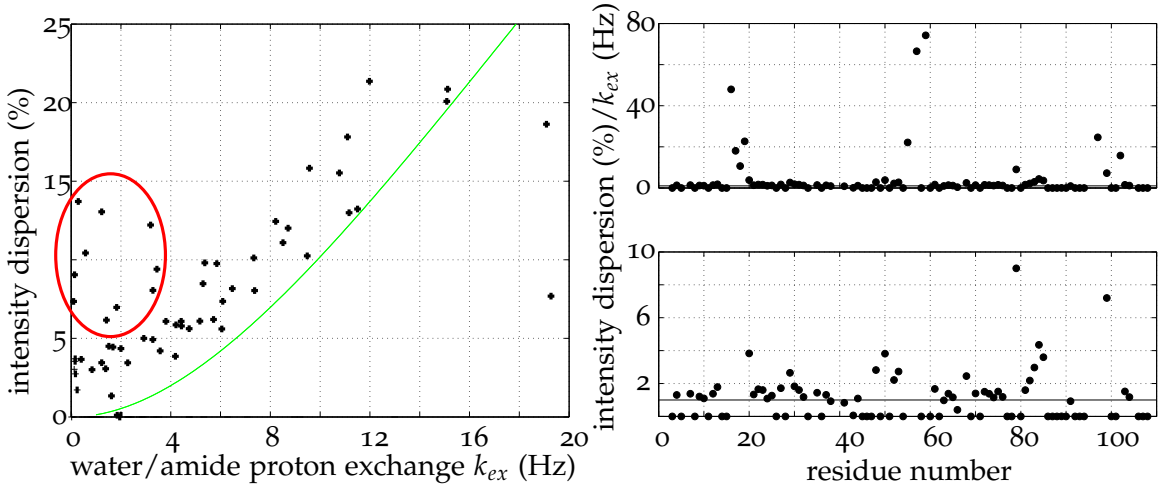


Figure 8.23 – *Left : relaxation dispersion as a function of water/amide proton exchange rates for CoblAB (in black). The green curves shows the expected contribution of water/amide proton exchange to relaxation dispersion for a non-optimized relaxation dispersion pulse sequence. Right : percentage of intensity dispersion divided by proton amide exchange rates as a function of protein sequence highlights regions with conformational exchange in a micro to millisecond time range.*

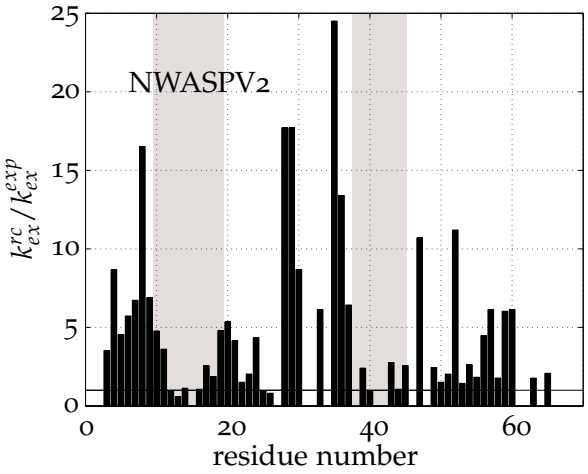


Figure 8.24 – *Protection factors of NWASPV2 at 278K at pH=7.*

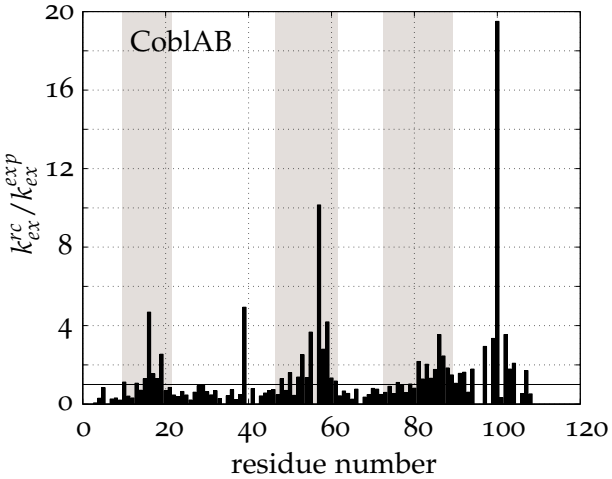


Figure 8.25 – *Protection factors of CoblAB at 278K at pH=7.*

Protection factors for NWASPV2 are shown on figure 8.24. Overall, protection factors values are around or below 1 except in the second halves of the three residual helices, in the C-terminal Streptag region and for Arg39 in i-1 of Pro40. Therefore the three residual helices do not have homogeneous dynamics in their whole sequences. The N-terminal regions of these three helices Ser10 to Leu15, Glu47 to Arg52 and Ser73 to Gln80 respectively are in exchange with random coil conformations at rates higher than the millisecond. The C-terminal region of these helices are more stable and the second and third helices may be stabilized by N-capping by Ser53 and Ser81 respectively.

8.4 Conclusion on the analysis of free WH2 repeats

The aim of this part of the project was to characterize the structural and dynamic behavior of different free WH2 repeats. These domains are known to be intrinsically disordered in absence of actin. Propensity to form the amphipathic N-terminal helix was observed for βT domains (Domanski et al. 2004, Cantrelle 2010). But until recently, there was little quantitative description of the conformational ensembles of WH2 repeats (Elazari-Shalom et al. 2015).

Here we analyzed the conformational ensembles of 6 WH2 repeats in 4 protein fragments.

CH1 and CH2 are βT domains chimeras of thymosin β4 and of the first domain of Ciboulot. Their sequences are almost identical except in the linker region between the amphipathic α-helix and the LKKT motif but they have different functions of promotion of unidirectional actin assembly and monomer sequestration respectively. In the actin-bound form, they both fold into a three-turn amphipathic α-helix followed by a 3-10 helix turn in the linker region, a β-strand and a second C-terminal α-helix (see section 3.3.1).

N-WASP comprises two short WH2 repeats of highly similar sequences which bind two actin monomers *in vitro*. These domains fold into two amphipathic helices followed by less defined, probably extended regions (see section 3.4).

The first two WH2 repeats of cordon-bleu were shown to interact with one actin in the 10^{-7}M range and two actins in the 10^{-5}M range. These two WH2 repeats have more different sequences and fold into two amphipathic α-helices and a third C-terminal α-helix. The second amphipathic α-helix interacts with two actin monomer on a CoblAB:actin$_2$ crystal structure (see section 3.5).

The most conserved structural feature between βT/WH2 repeats is the N-terminal amphipathic helix which interacts in the hydrophobic cleft between subdomains 1 and 3 of actin. Various predictors (AGADIR, S2D) find helical propensities in these regions.

All six amphipathic α-helices are partially formed in solution with variable populations : 10 to 15% for CH1, CH2 and the first domain of NWASPV2, $\simeq$25% for the second domain of NWASPV2 and the first domain of CoblAB and 40-45% for the second domain of CoblAB.

Experimental helical propensities in free WH2 repeats are not correlated with actin-binding affinity. For example, the equilibrium dissociation constant for CH2 is 250 times smaller than for CH1 but the two proteins have similar helical propensities. Actin affinity for mostly unfolded CH2 is in the nanomolar range whereas affinity for CoblAB for which amphipathic helices are two to three times more populated is in the 100-500nM range.

Linkers of βT peptides transiently form a 3-10 helix turn even in absence of actin. According to RDC experiments, this helix is more populated in CH2 than is CH1 which is consistent with stabilization by a salt bridge between Asp24 and Lys27.

The C-terminal region of βT petides which folds into a α-helix in the actin bound form remains disordered in the free peptides indicating that interaction with residues on the actin surface are necessary for the helix to fold. Positively charged side chain of Lys42 of CH1/CH2 interacts with Glu244 of actin, negatively charged side chain of Glu48 of CH1/CH2 interacts with Arg62 of actin.

A transient 4.5-turns α-helix was identified in the C-terminal region of domain B in CoblAB which is longer than the 2-turns α-helix observed in the crystal structure of actin-bound CoblAB. There are two possibilities to explain these differences : the crystal structure does not show the full C-terminal helix because of dynamic interactions or there is a partial unfolding process upon binding. Interactions of C-terminal regions of WH2 repeats are often more dynamic than the N-terminal ones which makes them more complicated to characterize by X-ray crystallography. Thus it would be interesting to characterize CoblAB-actin interactions in solution to confirm the size of this helix in the actin-bound form.

Experimental helical propensities were compared with values predicted by AGADIR (Munoz and Serrano 1994). This predictor only takes into account local helical propensity and not long range tertiary contacts that may stabilize interacting helices.

AGARDIR indeed predicts residual helicity in regions that form helices in the actin-bound form. However, it fails to predict absolute helix populations as well as relative populations between different proteins or between different helices of a same construct. Thus is is not possible to extrapolate directly helical propensities for other WH2 repeats.

Chimera of Tβ4 and CibD1 are mainly unfolded and their dynamics is dominated by fast backbone motions on the pico-nanosecond timescale that can be modeled by a dominant crankshaft transition thus helical and random coil conformers appear to be in fast equilibrium.

In NWASPV2, we identified additional long range interactions between the two amphipatic helices of NWASPV2 thanks to RDCs and relaxation dispersion experiments. No

Protein	Secondary structure	Experimental	AGADIR
CH1	N-ter amphipathic α-helix (14-22)	10-15%	2.7%
	linker 3-10 α-helix (24-27)	19%	2.1%
	C-ter α-helix (41-54)	3%	1.2%
CH2	N-ter amphipathic α-helix (14-22)	10-15%	2.8%
	linker 3-10 α-helix (24-27)	25%	3.3%
	C-ter α-helix (41-54)	3%	1.2%
NWASPV2	1st amphipathic α-helix (10-19)	15%	13.5%
	2nd amphipathic α-helix (38-47)	31%	12%
CoblAB	1st amphipathic α-helix (10-21)	25%	1.8%
	2nd amphipathic α-helix (47-61)	42%	34.2%
	C-ter α-helix (73-81)	20%	2.8%

Table 8.2 – *Comparison of experimental and predicted populations of conformers with folded amphipathic helices.*

similar interactions were detected in the case of CoblAB. These interactions may participate in the stabilization of transient amphipathic helices but it is hard to say whether they are biologically relevant. In the active state, the VCA extensions of nucleation promoting factors involved in the activation of the Arp2/3 complex are expected to be extended to allow G-actin monomers recruitment by WH2 repeats to promote assembly of a new daughter filament. However, in the inactive state, they do not bind actin so their WH2 repeats are not accessible. The crystal structure of the auto-inhibited WAVE1 complex (Chen et al. 2010) shows that its WH2 repeats folds into a α-helix which interacts with the Sra1 subunit (see figure 3.14). In WASP/N-WASP, no such interactions were identified and the WH2 repeats are expected to remain disordered. Maybe transient interactions between the two WH2 repeats of N-WASP could help to stabilize conformations where they are not accessible to actin.

In CoblAB, helical conformers are more populated, especially in domain B. These secondary structures remain stable at higher temperature (298K). Thus contribution of fast pico to nanosecond dynamics is reduced in these regions. Additionally, some residues display important chemical exchange at the C-termini of the two amphipathic helices in the micro to millisecond range.

Free WH2 repeats have diverse behaviors in solution, from a 90% random coil polymer-like peptide chain to 40% folded helical conformers in equilibrium with unfolded states (see figure 8.26). From the characterization of free WH2 repeats, it is difficult to draw conclusions on their function : CH1 and CH2 have different functions but similar structural and dynamic behavior. It is also complicated to predict the structure of actin-bound repeats. Possible C-terminal helices are either disordered (CH1 and CH2 or transiently populated (CoblAB). The only common structural feature between domains is the tran-

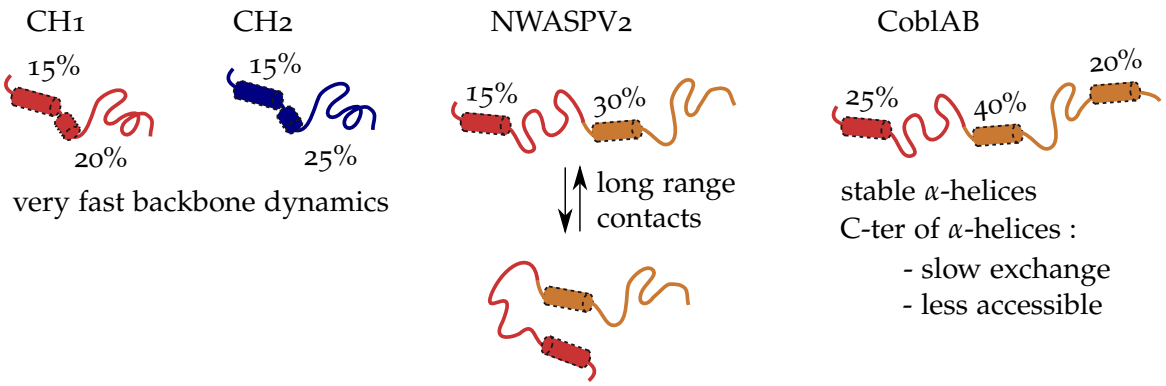


Figure 8.26 – *Residual structure and dynamics in free WH2 domains.*

siently populated amphipathic α-helix. This amphipathic helix appears to be an important anchor point in a possible conformational selection mechanism of interaction with actin. The disordered C-terminal regions of chimeras indicate there might also be contributions of induced fit in the interaction.

Besides, this *in vitro* characterization is only a partial view of the picture, NWASPV2 and CoblAB are small modules in large multi-domain proteins and structural propensities will probably be influenced by adjacent domains and free WH2 repeats may be involved in autoinhibition mechanisms like in the WAVE protein.

Mechanisms of actin interaction 9

This chapter describes actin:WH2 interaction experiments for CH1, NWASPV2 and CoblAB. Except when otherwise indicated, actin was recombinant wild-type drosophila 5C-actin overexpressed in insect cells.

9.1 CH1:actin binding as a function of salt concentration

We analyzed folding upon actin binding for single WH2 repeats.

As explained in chapter 3, interactions between βT/WH2 domains and actin depend on ionic strength and KCl concentration. In G-buffer, in absence of salt, CH1 and CH2 are tightly bound to actin via an amphipatic N-terminal α-helix which interacts with the hydrophobic cleft between subdomains 1 and 3 of actin followed by a 3-10 helix turn in the linker region a β-strand and a C-terminal α-helix which interacts with the pointed end of actin. Increase of KCl concentration leads to release of the C-terminal α-helix and then release of the whole peptide at high ionic strength for CH1. CH2 has a different behavior and no difference of dynamics was observed between the N-terminal and C-terminal regions of the protein because of the stabilization of the central and C-terminal regions by electrostatic interactions involving the linker regions (Didry et al. 2012).

Exchange between the free and bound CH1 and CH2 was followed by relaxation dispersion and chemical exchange saturation transfer experiments in order to probe possible intermediate states in folding upon binding.

^{15}N-relaxation dispersion experiments were recorded on ^{15}N-labeled CH1 in presence of actin. Previous Nzz exchange experiments showed that in G-buffer, at 303K, exchange rate beween free and bound CH1 was about 14Hz (Cantrelle 2010), which is low for relaxation dispersion experiments. However, relaxation dispersion also depends on the population of the minor state and chemical shift differences. Therefore we were counting on important chemical shift differences linked to folding upon binding processes as well tunable population on the minor bound state to optimize relaxation dispersion profiles. Various experimental conditions were tested but no exchange was detected. A possible reason for that is that exchange between the free and bound CH1 may be indeed too slow for relaxation dispersion experiments which are optimal for exchange rates k_{ex} between 100 and 10000Hz.

Chemical exchange saturation transfer experiments allow detection of slower exchange processes with exchange rates ranging from 20 to 200Hz. The recently published pulse

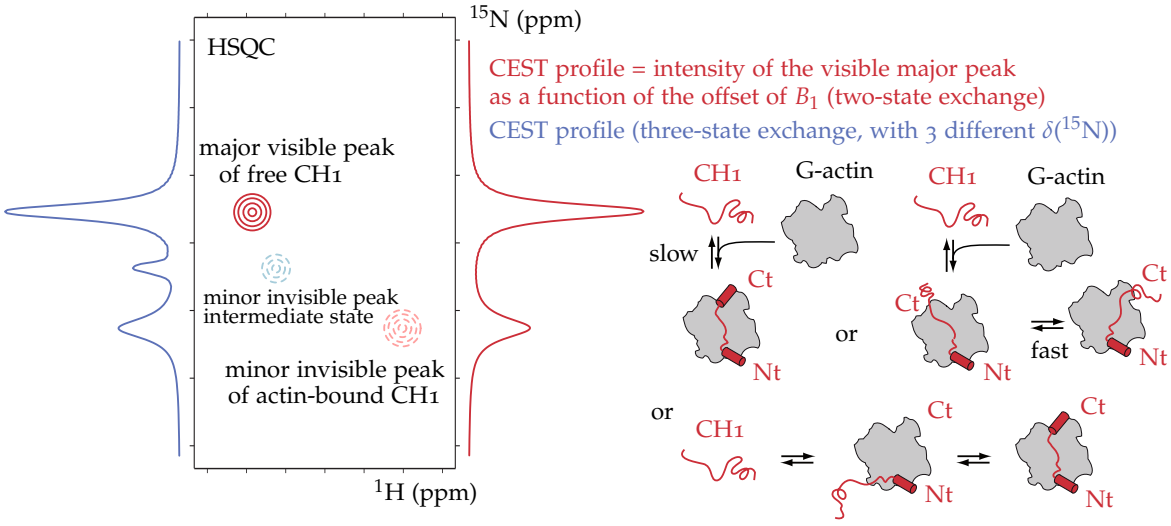


Figure 9.1 – *CEST experiments to analyze CH1 free/bound exchange. CEST profiles for every residues are plotted intensity of the major, visible peak on the HSQC spectrum, as a function of the offset of the B_1 field. If this major state is in slow exchange with one or several states with different chemical shifts, intensity dips will be observed on the CEST profiles for corresponding B_1 field offsets. Here, we are investigating exchange between free and actin-bound CH1 and different possibilities are : a. two-state exchange, b. three-states exchange with a fast exchange regime between two of the three states and c. three-states exchange with slow exchange regime between the three species of different chemical shifts.*

sequences of ^{15}N-CEST experiments (Vallurupalli et al. 2012) were implemented on the spectrometers and validated on a model protein (see experimental section for more details).

^{15}N-CEST experiments were recorded on free CH1 and CH2 in G-buffer, pH=7 at 298K and in absence of actin and no exchange was detected.

^{15}N-CEST experiments were recorded on CH2 in presence of actin ($\simeq$ 5% molar), at 298K, in G-buffer (no salt) and in F-buffer (150mM KCl) and no exchange was detected in agreement with very slow exchange kinetics measured by z-z exchange experiments (Cantrelle 2010) : the lifetime of (CH2:actin) complex is superior to 1 second in G-buffer at 303K.

^{15}N-CEST experiments were recorded on the same ^{15}N-labeled CH1 sample in presence of actin ($\simeq$ 5% molar), at 298K, in G-buffer (omM KCl) and with increasing KCl concentrations (80, 100, 125 and 150mM KCl). For each experimental condition, two sets of data were acquired with B_1 field frequencies of 12.5 and 25Hz. The sample showed significant evolution during the experiments at 150mM KCl (general decrease in the peak intensities) so the corresponding data were not used for the analysis. ^{15}N-CEST profiles show there is exchange between the major free CH1 and a minor state in G-buffer and that the exchange parameters depend on KCl concentration (see figure 9.2 for a few examples). The minor intensity dip becomes more important as the concentration in potassium chloride increases. Actin concentration is constant in all experiments so this change in intensity dips is consistent with a global increase of exchange rate between the free and bound

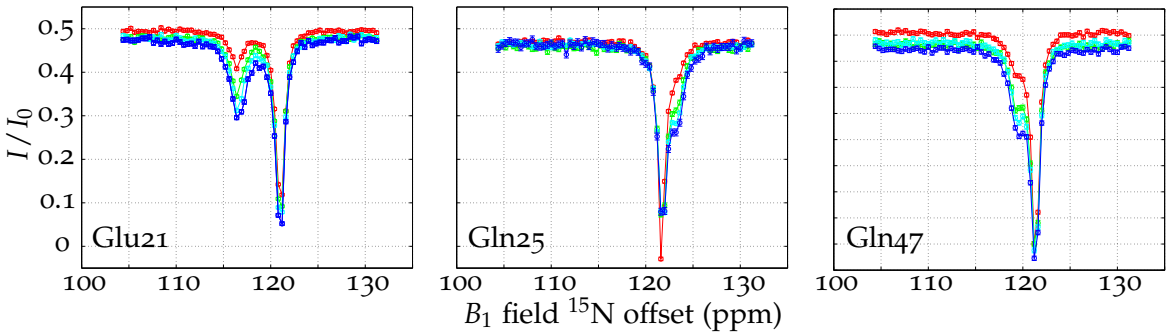


Figure 9.2 – *CEST profiles of Glu21, Gln25 and Gln47 of CH1 in G-buffer, omM KCl (red), 80mM KCl (green), 100mM KCl (cyan) and 125mM KCl (blue).*

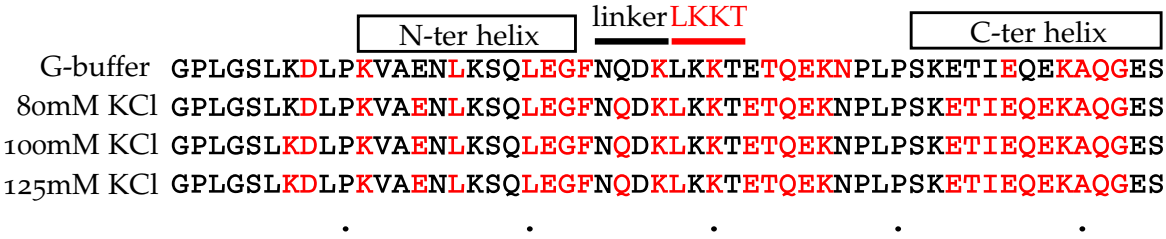


Figure 9.3 – *Residues used in the adjustment of CEST curves for the different buffer conditions.*

forms, in agreement with different equilibrium dissociation constants for the (CH1:actin) complex ($K_D = 0.25\mu$M in G-buffer and $K_D = 3.4\mu$M in F-buffer) derived from AEDANS fluorescence titration curves.

The CEST profiles depend on the exchange parameters p_B, k_{ex}, chemical shift of the minor state δ_B but also on the relaxation parameters of the free and bound form R_1, R_2^B and R_2^F. Because of the quantity of parameters to adjust, we did not perform profiles fitting on a per residue basis but adjusted exchange parameters for clusters or regions of the protein, assuming a global exchange model with a single p_B and a single k_{ex} for the cluster.

For each KCl concentration, CEST profiles showing two clearly separated intensity dips were first fitted to a global two state exchange model. CEST data of 20, 27, 28 and 28 residues (shown on figure 9.3) for experiments in G-buffer, 80mM KCl, 100mM KCl and 125mM KCl respectively were used for the global fit. The minor state population was fixed at 4.5% because actin concentration was the same for the different sets of experiments. The adjusted exchange rates were $k_{ex} = 8.99 \pm 0.24$Hz in G-buffer, $k_{ex} = 17.18 \pm 0.24$Hz in 80mM KCl, $k_{ex} = 22.46 \pm 0.26$Hz in 100mM KCl and $k_{ex} = 27.18 \pm 0.28$Hz in 125mM KCl. Chemical shifts of the minor population do not vary with KCl concentration and correspond to the chemical shifts of actin-bound CH1.

CEST profiles were then fitted using a variable p_B to a global two state exchange model, for the whole domains and for regions of the protein : the N-terminal amphipathic helix (residues 11 to 23), the linker, LKKT region and central regions (residues 24 to 38) and the C-terminal helix (residues 41 to 54). Results are summarized on figure 9.4.

The exchange rate k_{ex} for the global fit on the whole protein increases with KCl concen-

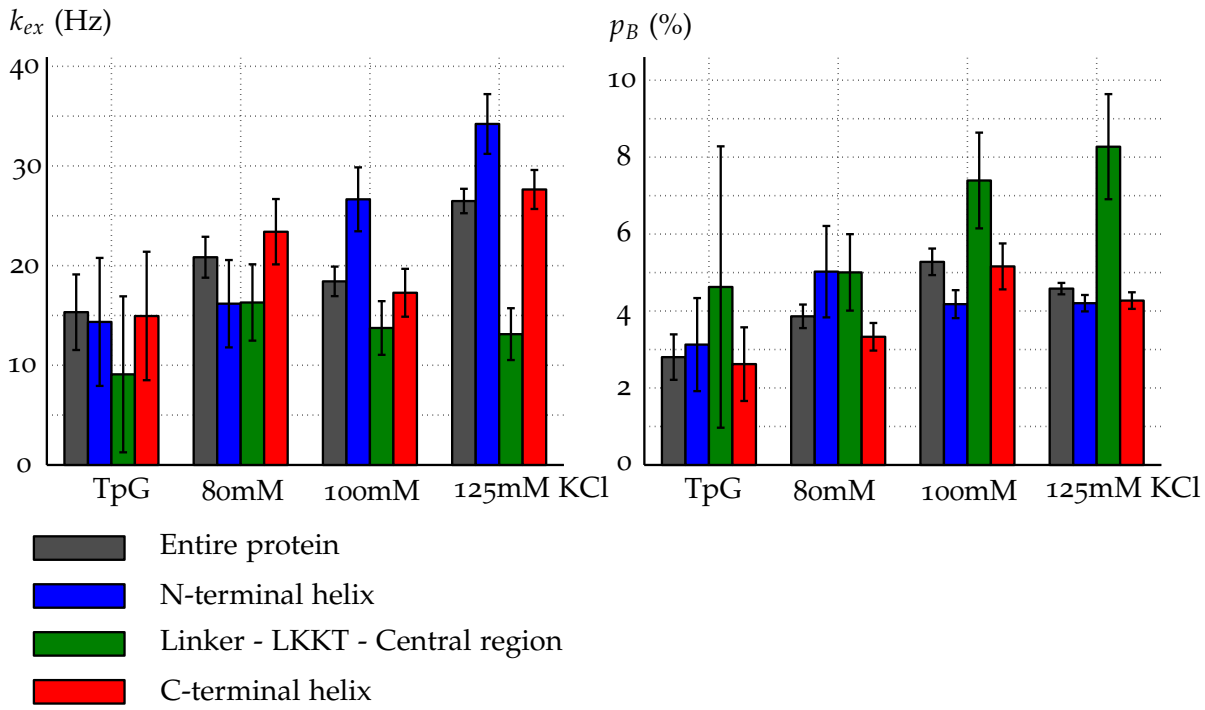


Figure 9.4 – *Conformational exchange parameters obtained by adjustment of CEST data, for the full protein, for the N-terminal amphipathic α-helix, for the linker, LKKT and central regions and for the C-terminal helix. Only residues showing two clearly separated intensity dips in their CEST profiles where used for exchange parameters determination.*

tration. Exchange rates and populations of the minor state vary depending on the protein region suggesting that this might be more complicated than a global two-state exchange process.

In G-buffer, the k_{ex} is similar for the global fit on the whole protein, the N-terminal amphipathic α-helix and the C-terminal α-helix. k_{ex} in the central region is lower but there is also more uncertainty in its determination due to poorer data quality in this region. Overall, CEST data in G-buffer are in agreement with a global two-state exchange model.

At 80mM KCl, the exchange rate is similar in the N-terminal helix and in the central region and it is higher in the C-terminal region. These results support previous finding on the release of the C-terminal α-helix of CH1 when the ionic strength increases.

At 100mM KCl and 125mM KCl, the exchange rate is lower for the central region and higher in the N and C-terminal helices. It seems to indicate that after the release of the C-terminal helix, interaction of the N-terminal α-helix becomes more dynamic as well. At these KCl concentrations, according to the fit, the central region of CH1 may be the main anchor point in the interaction. The adjusted populations for the minor state in the central region is higher than expected, since the estimated CH1:actin molar ratio is 1:0.05. A first possibility to explain this difference is the important uncertainties on the protein concentration of the sample. CH1 has little aromatic residues, only one phenylalanine, and so does not absorb at 280nm. Concentration is determined by a BCA assay and may

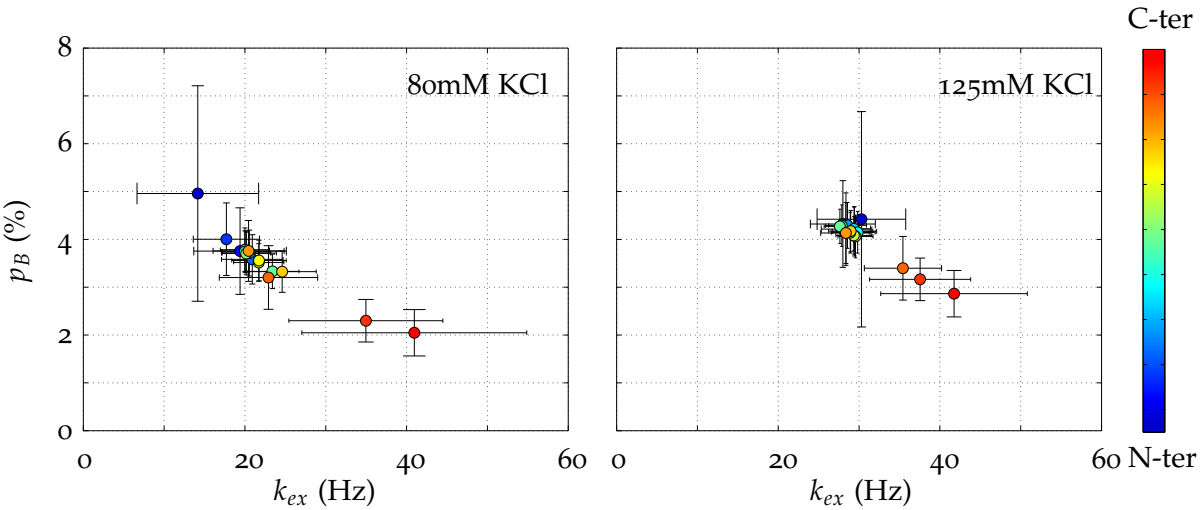


Figure 9.5 – *Conformational exchange parameters obtained by adjustment of CEST data for clusters of residues in the C-terminal helix. The color code is the following : blue dots are for clusters of residues at the N-terminus of the α-helix and red dots are for clusters of residues at the C-terminus*

be source of errors. Another possibility is a similar dependence on p_B and k_{ex} of the CEST profiles (see figure 6.11). To check for possible correlation between p_B and k_{ex}

A particular behavior was observed in the C-terminal helix at 80mM KCl. CEST data of residues in the C-terminal helix was analyzed by fitting different clusters of pics at different position in the helix : residues (43; 44), (43; 44;45)... (43; 44; 45; 46; 47; 48; 49; 50; 51; 52)... (51; 52). Results are displayed on the left panel figure 9.5. Lower k_{ex} and higher p_B values were found for cluster with residues at the N-terminus of the helix, in blue on figure 9.5. And higher k_{ex} and lower p_B values were found for cluster with residues at the N-terminus of the helix, in blue on figure 9.5. These results seem to indicate that at 80mM KCl, the C-terminal helix is not completely released and interaction with actin is stronger for the first residues of the helix.

At 125mM KCl, there is no such phenomenon. For most clusters of residues of the C-terminal helix, similar exchange rates and minor state populations were found, around 28Hz and 4% respectively (right panel of figure 9.5). Only clusters containing the few last residues of the helix displayed faster exchange rates between 35 and 45Hz. These result are compatible with a more dynamic interaction between the full C-terminal helix and actin. A similar behavior is expected in G-buffer when the C-terminal helix is tightly bound but we were not able to verify by experiment because only 5 residues of the C-terminal helix (Glu46, Lys49, Gln50, Ala51 and Gly52) had CEST profiles with two clearly separated intensity dips.

We were able to follow exchange between free and bound CH1 by ^{15}N-CEST experiments in different buffer conditions. CEST was previously used to follow exchange between three state with different chemical shifts (Lim et al. 2014). In that case, there were three intensity dips in CEST profiles. Here, a maximum of two intensity dips were ob-

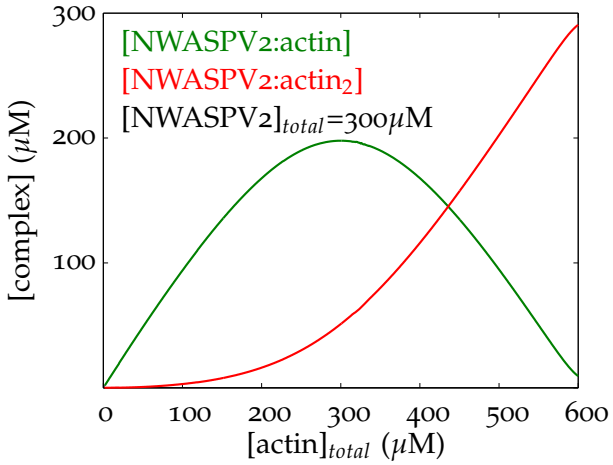


Figure 9.6 – *Expected concentrations of 1:1 and 1:2 complexes as a function of actin concentration in the case of a non cooperative interaction. Curves were simulated for a constant total NWASPV2 concentration of 300μM and dissociation constants of 20 and 300nM.*

served in CEST profiles for all residues whose chemical shifts correspond to those of the free and bound CH1. So we did not detect any third intermediate state with different chemical shifts from the free and bound forms. As we saw in chapter 8, the N-terminal helix of CH1 is preformed in solution whereas the C-terminus is completely disordered. It is therefore unlikely that CH1 binds actin via a pure conformational selection mechanism but we still have no information on a possible intermediate state.

For KCl concentrations above 80mM, CEST data suggests that the exchange process is not a simple two-state exchange model. CEST data indicate that the C-terminal α-helix becomes more dynamic for KCl concentrations above 80mM KCl and that the N-terminal helix becomes more dynamic for KCl concentrations above 125mM. The region which is less affected by ionic strength is the center of the peptide including the LKKT consensus motif implying that this region is a major anchor point in the interaction.

9.2 Interaction of NWASPV2 with actin

NWASP V2 is expected to bind two actin monomers *in vitro* with affinity constants of 20 and 300nM (Gaucher et al. 2012). Fluorescence anisotropy of actin-AEDANS showed that both NWASPV2:actin and NWASPV2:actin$_2$ complexes were present in solution depending on NWASPV2:actin ratio.

The expected concentrations of NWASPV2:actin and NWASPV2:actin$_2$ complexes as a function of actin concentration, assuming a non-cooperative interaction, are shown on figure 9.6. In our titration experiments by NMR, the total NWASPV2 concentration was kept constant (300μM). In the hypothesis of a non-cooperative mechanism, 1:1 complexes are preferentially formed for total actin concentrations below 300μM and 1:2 complexes become predominant for total actin concentrations above 450μM.

Interaction of NWASPV2 with actin was followed by NMR spectroscopy. Increasing

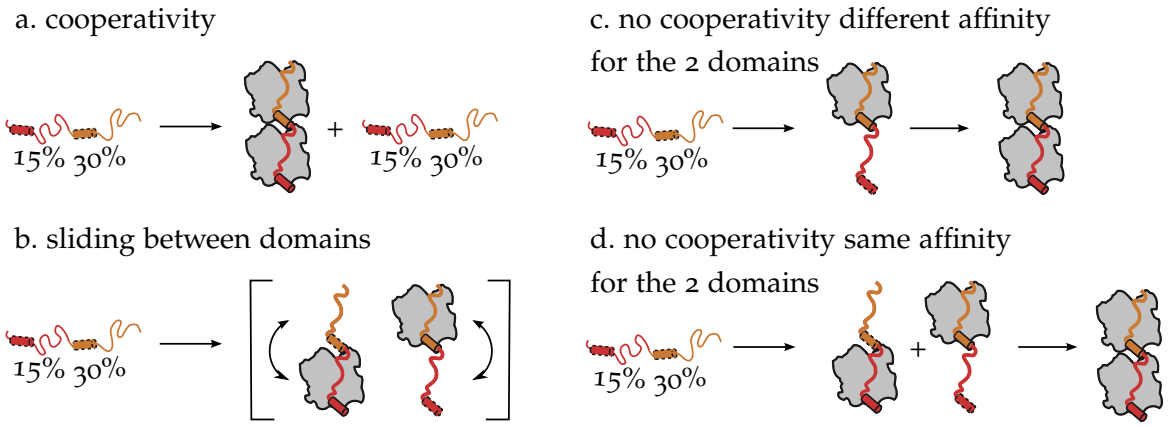


Figure 9.7 – *Possible interaction modes for a WH2 bidomain and two actin monomers (see text for details).*

amounts of unlabeled actin were added to a 300μM solution of ^{15}N-labeled NWASPV2, in G-buffer, pH=7, at 278K and 950MHz. (^{1}H-^{15}N)-HSQC spectra recorded at different NWASPV2:actin ratios are shown on figure 9.8. Assuming a non-cooperative interaction and binding affinities determined by Gaucher et al. (2012), expected concentrations of NWASPV2:actin and NWASPV2:actin$_2$ are $\simeq$15μM and $\simeq$135μM for a 1:0.5 NWASPV2:actin concentration ratio, $\simeq$50μM and $\simeq$200μM for a 1:1 NWASPV2:actin concentration ratio and $\simeq$150μM and $\simeq$150μM for a 1:0.5 NWASPV2:actin concentration ratio.

Peak intensities decreases upon actin addition and not significant peak shifts are observed. This is consistent with a slow exchange regime as previously observed for βT/WH2 repeats of Tβ4, CibD1, their chimeras and WASP (Domanski et al. 2004, Hertzog et al. 2004, Kelly et al. 2006, Didry et al. 2012).

New peaks are only observed in the lateral chains region of the spectrum. This is surprising since at 1:0.5 and 1:1 NWASPV2:actin ratios, we expected that the main complex in solution would have a 1:1 stoechiometry and a $\simeq$50kDa molecular weight. CH2:5C-actin complex has a similar molecular weight and the actin-bound CH2 was observable on HSQC spectra with no need to deuterate the peptide.

Most plausible explanations for this absence of peaks for the bound-form are intermediate exchange regime leading to dramatic line broadening of the peaks or cooperative binding and predominance of the 1:2 NWASPV2:actin complex even at low actin concentration. The 1:2 complex is over 100kDa therefore and direct observation of an object of this size by NMR spectroscopy requires at least deuteration of the N-WASP fragment. Another possibility is that actin binds to either WH2 repeat of NWASPV2 with the same affinity and two 1:1 complexes with a twofold lower concentration coexist in solution, too diluted to be detected by NMR.

Figure 9.9 shows intensity changes upon actin addition. Intensities in regions expected to bind actin most tightly (amphipathic α-helices and L++T/V motif) decrease similarly in the two domains, even at a 1:0.5 NWASP:actin concentration ratio where a 1:1 complex was expected to be largely predominant. Average intensities in these regions are summarized in table 9.1. Residues 3 to 7, 27 to 29 and 35 to 36 show more important intensity decrease

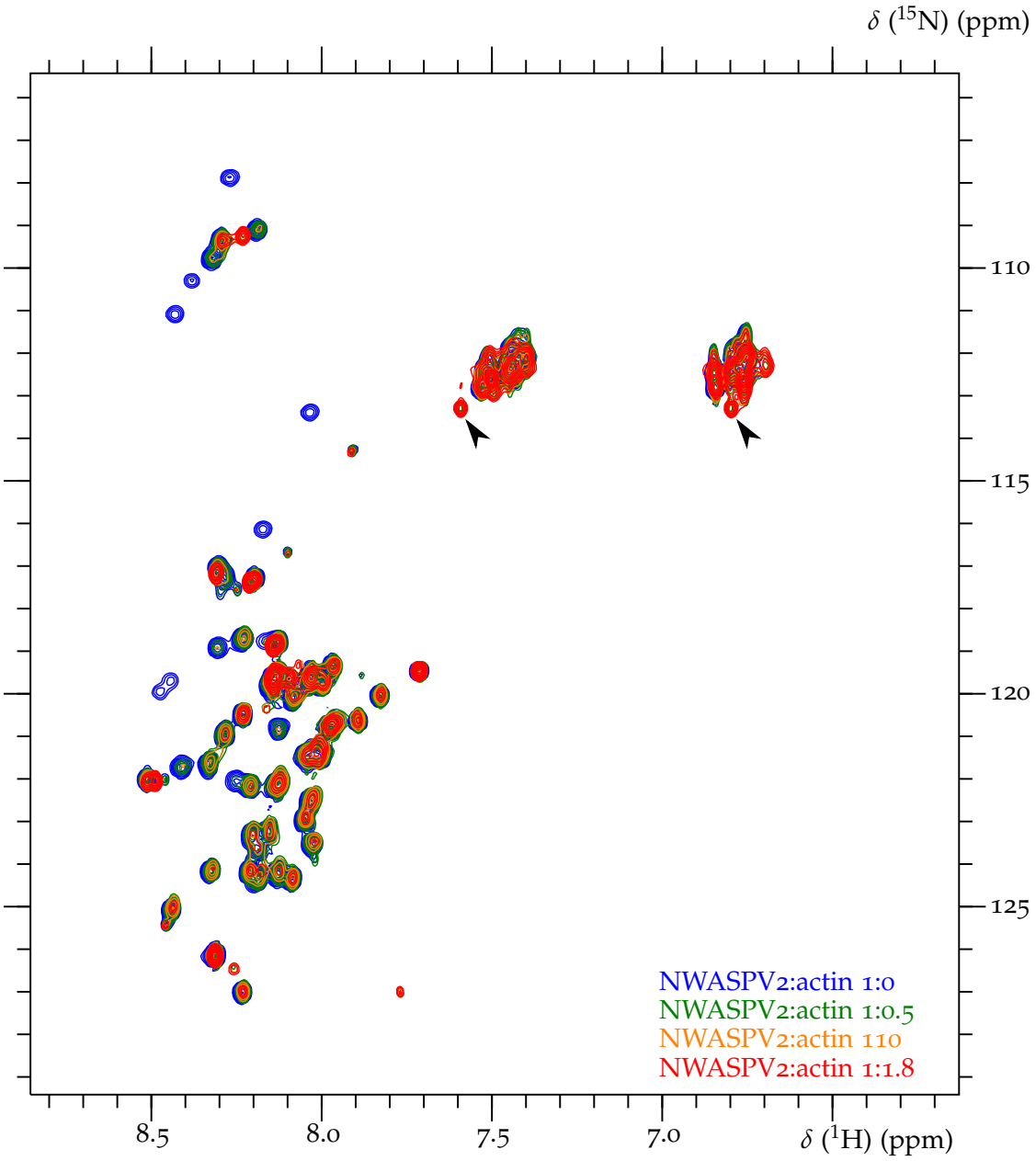


Figure 9.8 – *Superimposition of (^{1}H-^{15}N)-HSQC spectra of the WH2 bidomain of human N-WASP in presence of different amount of 5C-actin. Spectra were recorded at 950MHz, 298K and pH=7. The two peak which appear upon actin addition are indicated by black arrows.*

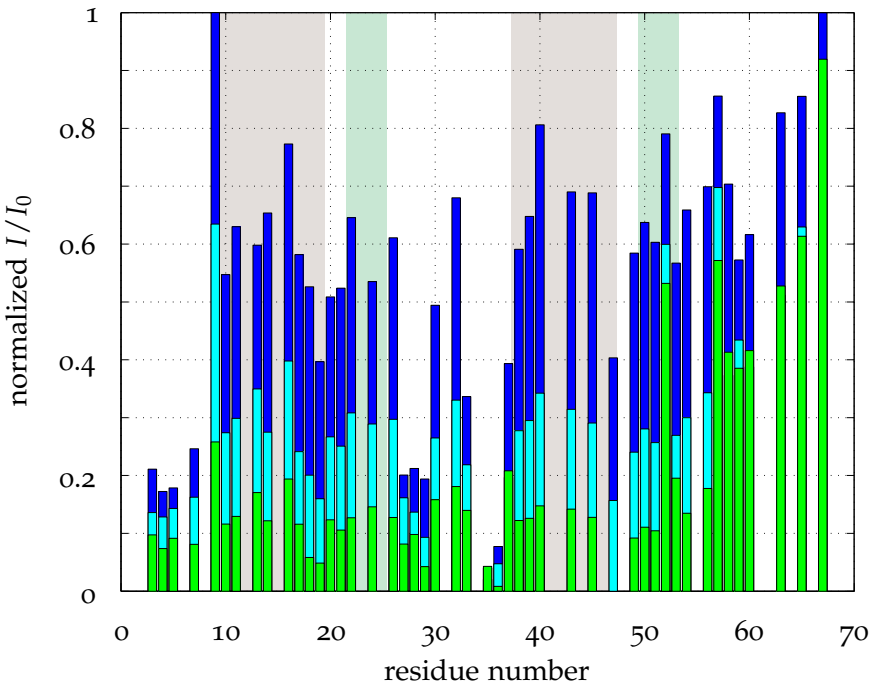


Figure 9.9 – *Normalized (^{1}H-^{15}N) HSQC peak intensities changes of free NWASPV2 in presence of actin. Blue, cyan and green bars correspond to 1:0.5, 1:1 and 1:1.5 NWASPV2:actin ratios respectively. Reference intensities are for (^{1}H-^{15}N) HSQC peaks on a free NWASPV2 sample. A correction factor was applied to account for NWASPV2 concentration differences between samples.*

	1st domain		2nd domain	
NWASP:actin	amphipathic α-helix	LKKV	amphipathic α helix	LKSV
1:0.5	59%	59%	64%	65%
1:1	27%	30%	30%	35%
1:1.5	11%	14%	13%	24%

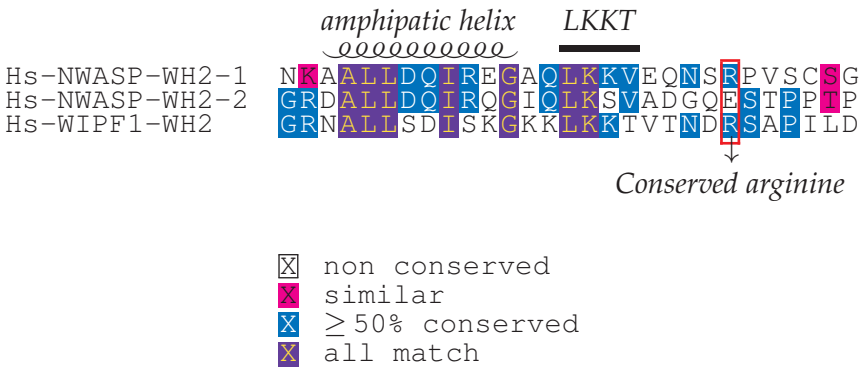
Table 9.1 – *Normalized (^{1}H-^{15}N) HSQC peak intensities of NWASPV2 in presence of actin (0.5 molar equivalent) compared to intensities in absence of actin.*

even for low actin concentrations. This is unexpected and hard to explain since these regions do not belong to particular actin binding region or motifs.

These results are not compatible with no cooperative binding with binding sites of different affinity but is rather consistent with a cooperative interaction between the WH2 tandem repeat of N-WASP and actin or actin binding to either WH2 repeat of N-WASP, non cooperative binding with the same affinity for the two sites or sliding of actin between the two sites. Gaucher et al. (2012) previously observed cooperative binding in gel-filtration experiments, but not in the actin-AEDANS fluorescence titration experiments to determine the binding affinity constants. It is possible that actin modification with AEDANS modifies NWASPV2 binding and abolishes cooperativity.

Relative peak intensities in the C-terminal region of the protein are higher than for the rest of the sequence. For a 1:1.5 NWASPV2:actin, average peak intensities are 55% of initial intensities for residues 56 to 67 and 17% of initial intensities for the rest of the

protein. Therefore, the last 12 residues of NWASPV2, including the proline-rich region at the C-terminus, do not bind actin as tightly as the rest of the protein and strong WH2:actin interactions only extends for two residues downstream the second L++T/V motif.



Sequence alignment of N-WASP and WIP WH2 repeats reveals that conserved arginine 54 of WIP, which forms a salt bridge with actin Glu93, is found in the first WH2 domain of N-WASP but not in the second where it is replaced by a glutamate. Disruption of this salt bridge might explain the weaker interaction of the C-terminal region of the second WH2 repeat of N-WASP with actin.

9.3 Interaction of CoblAB with actin

9.3.1 Titration experiments by NMR

CoblAB interacts with two actin monomers at high protein concentration *in vitro*. Equilibrium dissociation constants measured for CoblA, CoblB and for the first binding site of CoblAB are 0.1, 0.5 and 0.3μM respectively. At the low concentrations used in the fluorescence experiments to determine the dissociation constants, CoblAB only binds one actin but at higher concentrations, a 1:2 complex is formed.

We followed interaction of actin with CoblAB by NMR titration experiments. CoblAB was uniformly ^{15}N- and ^{13}C labeled and 80% deuterated. (^{1}H-^{15}N)-HSQC spectra were recorded in G-buffer pH=7, at 298K and 950MHz in absence and in presence of 1, 1.5 and 1.75 molar equivalent of actin (see figure 9.10). Peak intensities decrease as actin concentration increases which is in agreement which intermediate to slow exchange regimes.

As for NWASPV2, intensity changes upon actin addition were analyzed (see figure 9.11). At first sight, it seems that actin preferentially interacts with domain B of CoblAB. For a 1:1 NWASPV2:actin ratio, average I/I_0 values are 0.39 and 0.25 for the first and second WH2 domains respectively. But the average I/I_0 value in the first amphipathic helix is 0.3. Therefore, the entire domain B of CoblAB interacts with actin as well as the amphipathic helix of domain A, the central and C-terminal regions of domain A seem to remain flexible.

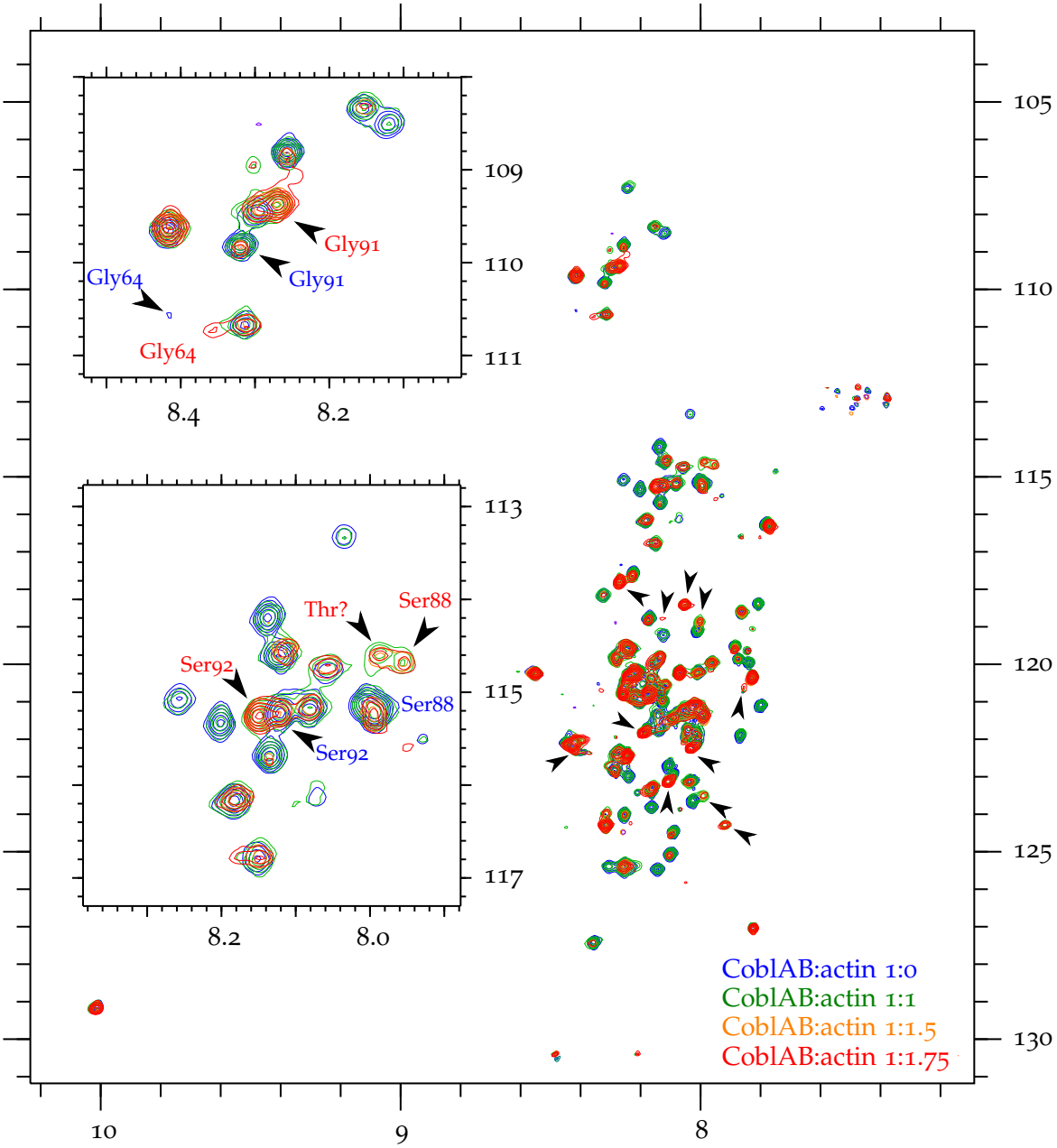


Figure 9.10 – *Superimposition of (^{1}H-^{15}N)-HSQC spectra of the WH2 bidomain AB of human cordon-bleu in presence of different amount of 5C-actin. Spectra were recorded at 950MHz, 298K and pH=7. Some peaks which appear upon actin addition are indicated by black arrows. Insets show the glycine and serine/threonine regions.*

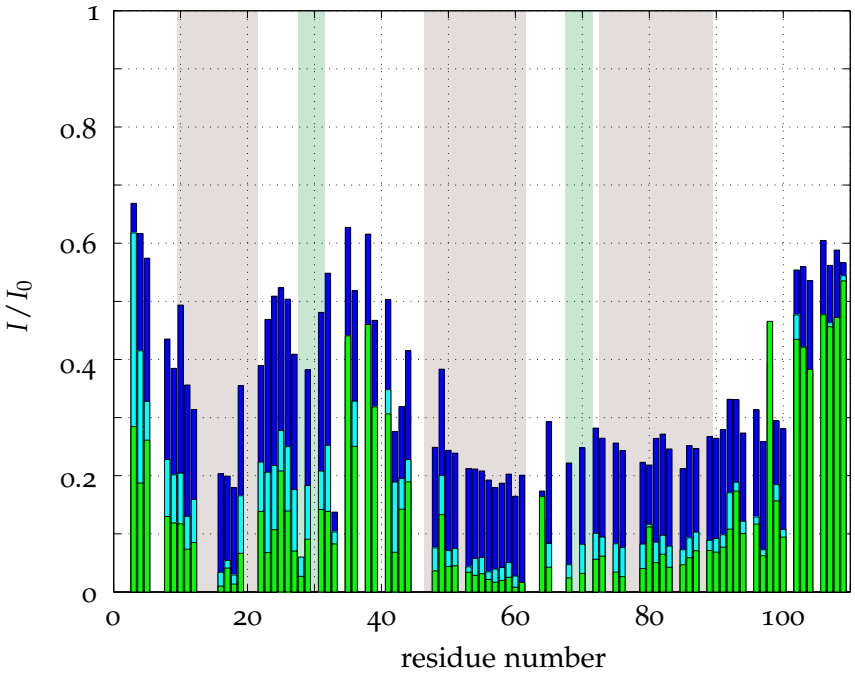


Figure 9.11 – *Normalized (^{1}H-^{15}N)-HSQC peak intensities changes of free CoblAB in presence of actin. Blue, cyan and green bars correspond to 1:1, 1:1.5 and 1:1.75 CoblAB:actin ratios respectively. Reference intensities are for (^{1}H-^{15}N) HSQC peaks on a free CoblAB sample. The two consensus LKKT motifs and three helices positions are highlighted in green and grey respectively.*

This long flexible region may serve the multifunctionality of Cobl to accommodate different (Cobl:actin$_2$) conformations involved in different steps of actin assembly regulation. But the exact role of this flexible linker still has to be investigated.

About 30 new peaks appear on the (^{1}H-^{15}N) HSQC spectra in presence of actin. Sequential and bidirectional three-dimensional experiments were recorded to assign these new peak : HNCO, HN(CA)CO, HNCACB and HN(CO)CACB. Eleven peaks were successfully assigned to the following residues : Gly64 (only ThrGly motif in the sequence), Ser88 to Ser94 and Leu94 to Leu100. Other peaks of lower intensity were not found on 3D experiments or sequential connections to other residues could not be established.

In the free CoblAB protein, residues 73 to 89 form a residual α helix with a population around 20%. In the crystal structure of actin-bound CoblAB (PDB : 4JHD), this helix is present but much shorter and stops at Ser81. The chemical shifts of the assigned new peaks were compared with those of the free form (see figure 9.12). There are very little chemical shift variations from Gly91 to Leu100. But for Leu87 to Gln90, $\delta_{bound} - \delta_{free}$ is positive for carbonyles and carbons α and negative for amide protons and nitrogens and carbons β which is consistent with a disorder to α-helix transition. Thus, chemical shifts analysis support the hypothesis of a long C-terminal α-helix in actin-bound CoblAB.

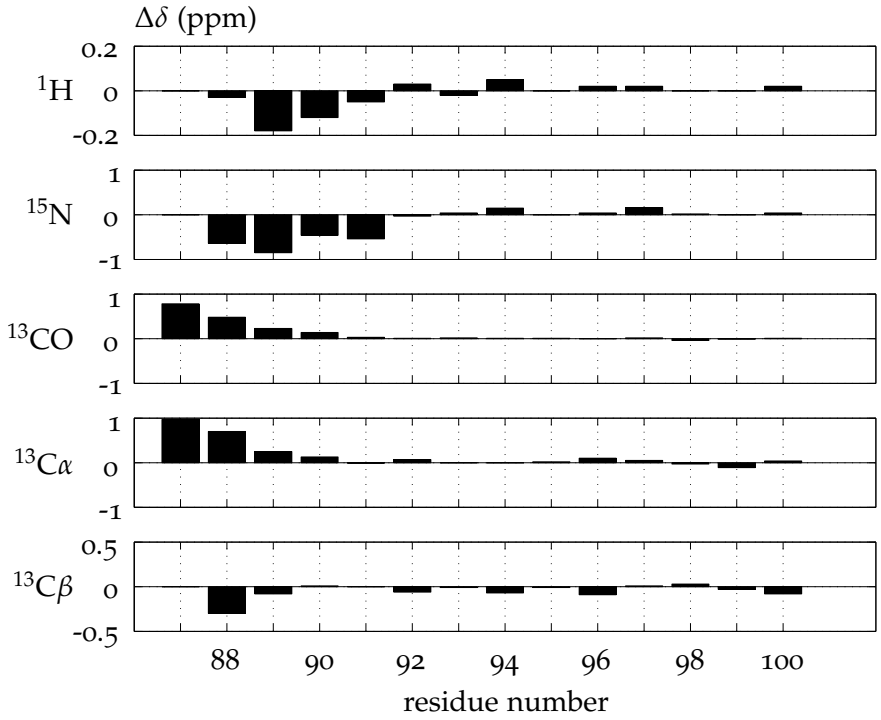


Figure 9.12 – *Chemical shift differences between the assigned free and bound peaks in the C-terminal region of CoblAB.*

General conclusion

This project was focused on understanding of the regulation of actin assembly by WH2 repeats. We set up production of recombinant actin for structural investigation. Characterization of the non-polymerizable actin mutant we planned to use for structural studies showed that it disrupts interactions with the C-terminal region of βT/WH2 repeats. We also showed that it was possible to label actin for NMR. Peaks on the (^{1}H-^{15}N)-HSQC spectra of ^{15}N-labeled actin are rather broad because of the high molecular weight of actin. Deuteration and ^{13}C-labeling will be necessary for further experiments and for assignment of at least some of the resonances.

We analyzed free WH2 repeats in absence of actin. We characterized residual secondary structure from chemical shifts analysis and residual dipolar coupling experiments and dynamics on timescales ranging from the pico to nanosecond for relaxation experiments to micro to millisecond for relaxation dispersion and CLEANEX experiments.

Our analysis of free WH2 repeats showed that it is difficult to draw general conclusions on residual structure and dynamics. The only common structural feature is the transiently formed N-terminal amphipathic α-helix. Population of these transiently formed helices is highly variable, from 15 to 45% in our experimental conditions.

WH2 bidomains showed more complex dynamics than single domains with long range interaction in the case of NWASV2 and slow conformational exchange in transient helices regions in the case of CoblAB. Paramagnetic relaxation enhancement experiments could help to confirm and characterize more precisely long range contacts in WH2 bidomains. Besides, it may be interesting to study CoblA and CoblB to check whether slow conformational exchange is linked to local sequence effects or enhanced by domain repetition. Finally, WH2 repeats are often part of large modular proteins as it is the case for NWASP and Cobl. It would be interesting to investigate the structural elements and dynamics of WH2 repeats in the context of the whole protein. Are there additional intramolecular interactions ? In absence of actin, are WH2 domains free and disordered in the whole protein or are they involved in autoinhibition mechanisms or interacting with other accessory proteins ?

Complexes between one or several actins and WH2 repeats are difficult to study. They display highly variable dynamics, and are often unstable *in vitro* as actin tends to polymerize. X-ray crystallography may provide information on structure and position of actin bound WH2 repeats but may not reflect the conformational diversity of these complexes.

In general, the C-termini of actin-bound WH2 repeats are not defined in crystal structures. In the case of CoblAB, the C-terminal helix seems to be longer in solution than of the crystal structure. Indeed crystal packing is susceptible to favor only some conformations.

NMR can provide information on structure and dynamics of complexes in solution. We followed interaction of WH2 bidomains with actin, we were able to identify interacting regions but had trouble to observe the complex. It is probably due to the high molecular weight of the complex and possible dynamics in the bound state leading to high relaxation rates.

A way to tackle this issue would be to use more sophisticated labeling schemes like methyl labeling of WH2 repeats. This may not be enough because in the bound form, actin protons participate to enhancement of relaxation rates. Therefore it might be necessary to deuterate actin to further improve relaxation properties of the sample. And this will require further development of isotopic labeling in insect cells.

Another solution is to not look directly to the complex but to characterize chemical exchange between free and bound states by looking at the free state. It was shown to be feasible by chemical exchange saturation transfer experiments on CH1:actin interaction and should be applicable to more complex interaction between WH2 bidomains and several actin. The main advantage of CEST experiments compared to for example Nzz exchange experiments is that they require only low populations of bound states enabling one to work with large excess of WH2 repeats in the samples and to limit risks of actin polymerization. A series of CEST experiment were acquired on CoblAB in presence of actin, no exchange was observe so far but experimental conditions will be optimized.

Bibliography

Abdul-Manan, N., Aghazadeh, B., Liu, G., Majumdar, A., Ouerfelli, O., Siminovitch, K., and Rosen, M. Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. *Nature*, 399(6734):379–383, 1999.

Ahuja, R., Pinyol, R., Reichenbach, N., Custer, L., Klingensmith, J., Kessels, M. M., and Qualmann, B. Cordon-bleu is an actin nucleation factor and controls neuronal morphology. *Cell*, 131(2):337–350, 2007.

Akkari, P., Nowak, K., Beckman, K., Walker, K., Schachat, F., and Laing, N. Production of human skeletal alpha-actin proteins by the baculovirus expression system. *Biochemical and Biophysical Research Communications*, 307(1):74–79, 2003.

Alexandrescu, A. and Shortle, D. Backbone Dynamics of a Highly Disordered 131-Residue Fragment of Staphylococcal Nuclease. *Journal of Molecular Biology*, 242(4):527–546, 1994.

Atkinson, R. and Kieffer, B. The role of protein motions in molecular recognition: insights from heteronuclear NMR relaxation measurements. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 44(3-4):141–187, 2004.

Bagshaw, C. *Muscle contraction*. Springer, 1992.

Bai, Y., Milne, J., Mayne, L., and Englander, S. Primary Structure Effects on Peptide Group Hydrogen-Exchange. *Proteins-Structure Function and Genetics*, 17(1):75–86, 1993.

Baldwin, A. J. and Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology*, 5(11):808–814, 2009.

Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638–1642, 2006.

Boivin, S., Kozak, S., and Meijers, R. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*, 91(2):192–206, 2013.

Bookwalter, C. and Trybus, K. Functional consequences of a mutation in an expressed human alpha-cardiac actin at a site implicated in familial hypertrophic cardiomyopathy. *Journal of Biological Chemistry*, 281(24):16777–16784, 2006.

Boomershine, W., Raj, M., Gopalan, V., and Foster, M. Preparation of uniformly labeled NMR samples in Escherichia coli under the tight control of the araBAD promoter: expression of an archaeal homolog of the RNase P Rpp29 protein. *Protein Expression and Purification*, 28(2):246–251, 2003.

Boquet, I., Boujemaa, R., Carlier, M., and Preat, T. Ciboulot regulates actin assembly during Drosophila brain metamorphosis. *Cell*, 102(6):797–808, 2000.

Borcherds, W., Theillet, F.-X., Katzer, A., Finzel, A., Mishall, K. M., Powell, A. T., Wu, H., Manieri, W., Dieterich, C., Selenko, P., Loewer, A., and Daughdrill, G. W. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nature Chemical Biology*, 10(12):1000+, 2014.

Bosch, M., Le, K. H. D., Bugyi, B., Correia, J. J., Renault, L., and Carlier, M.-F. Analysis of the function of spire in actin assembly and its synergy with formin and profilin. *Molecular Cell*, 28(4):555–568, 2007.

Bouvignies, G., Vallurupalli, P., Hansen, D. F., Correia, B. E., Lange, O., Bah, A., Vernon, R. M., Dahlquist, F. W., Baker, D., and Kay, L. E. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*, 477(7362):111–U134, 2011.

Brender, J. R., Krishnamoorthy, J., Sciacca, M. F. M., Vivekanandan, S., D'Urso, L., Chen, J., La Rosa, C., and Ramamoorthy, A. Probing the Sources of the Apparent Irreproducibility of Amyloid Formation: Drastic Changes in Kinetics and a Switch in Mechanism Due to Micelle like Oligomer Formation at Critical Concentrations of IAPP. *Journal of Physical Chemistry B*, 119(7):2886–2896, 2015.

Bubb, M., Govindasamy, L., Yarmola, E., Vorobiev, S., Almo, S., Somasundaram, T., Chapman, M., Agbandje-McKenna, M., and McKenna, R. Polylysine induces an antiparallel actin dimer that nucleates filament assembly - Crystal structure at 3.5-angstrom resolution. *Journal of Biological Chemistry*, 277(23):20999–21006, 2002.

Camilloni, C., De Simone, A., Vranken, W. F., and Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry*, 51(11):2224–2231, 2012.

Campellone, K. G. and Welch, M. D. A nucleator arms race: cellular control of actin assembly. *Nature Reviews Molecular Cell Biology*, 11(4):237–251, 2010.

Cantrelle, F.-X. *Étude par RMN de Protéines et de Peptides Régulateurs de la Polymérisation de l'Actine*. PhD thesis, Université Paris Sud, 2010.

Carlier, M.-F., Husson, C., Renault, L., and Didry, D. Control of Actin Assembly by the WH2 Domains and Their Multifunctional Tandem Repeats in Spire and Cordon-Bleu. In *International Review of Cell and Molecular Biology*, volume 290, pages 55–85. 2011.

Carlier, M. Actin - Protein-Structure and Filament Dynamics. *Journal of Biological Chemistry*, 266(1):1–4, 1991.

Carroll, E., Gerrelli, D., Gasca, S., Berg, E., Beier, D., Copp, A., and Klingensmith, J. Cordon-bleu is a conserved gene involved in neural tube formation. *Developmental Biology*, 262(1):16–31, 2003.

Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*, 420(6911):98–102, 2002.

Cavanagh, J., Fairbrother, W. J., III, A. G. P., Rance, M., and Skelton, N. J. *Protein NMR Spectroscopy*. Academic Press, Burlington, second edition edition, 2007.

Chen, C. K., Sawaya, M. R., Phillips, M. L., Reisler, E., and Quinlan, M. E. Multiple Forms of Spire-Actin Complexes and their Functional Consequences. *Journal of Biological Chemistry*, 287(13):10684–10692, 2012.

Chen, X., Ni, F., Tian, X., Kondrashkina, E., Wang, Q., and Ma, J. Structural Basis of Actin Filament Nucleation by Tandem W Domains. *Cell Reports*, 3(6):1910–1920, 2013.

Chen, Z., Borek, D., Padrick, S. B., Gomez, T. S., Metlagel, Z., Ismail, A. M., Umetani, J., Billadeau, D. D., Otwinowski, Z., and Rosen, M. K. Structure and control of the actin regulatory WAVE complex. *Nature*, 468(7323):533–U207, 2010.

Cheng, H.-C., Skehan, B. M., Campellone, K. G., Leong, J. M., and Rosen, M. K. Structural mechanism of WASP activation by the enterohaemorrhagic E-coli effector EspF(U). *Nature*, 454(7207):1009–U54, 2008.

Chereau, D., Kerff, F., Graceffa, P., Grabarek, Z., Langsetmo, K., and Dominguez, R. Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proceedings of the National Academy of Science of the United States of America*, 102(46):16644–16649, 2005.

Chesarone, M. A., DuPage, A. G., and Goode, B. L. Unleashing formins to remodel the actin and microtubule cytoskeletons. *Nature Reviews Molecular Cell Biology*, 11(1):62–74, 2010.

Cheung, M.-S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance*, 202(2):223–233, 2010.

Chhabra, E. S. and Higgs, H. N. INF2 is a WASP homology 2 motif-containing formin that severs actin filaments and accelerates both polymerization and depolymerization. *Journal of Biological Chemistry*, 281(36):26754–26767, 2006.

Chou, J., Gaemers, S., Howder, B., Louis, J., and Bax, A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of Biomolecular NMR*, 21(4):377–382, 2001.

Cino, E. A., Karttunen, M., and Choy, W.-Y. Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins. *PLoS ONE*, 7(11), 2012.

Clarkson, E., Costa, C., and Machesky, L. Congenital myopathies: diseases of the actin cytoskeleton. *Journal of Pathology*, 204(4):407–417, 2004.

Clore, G., Szabo, A., Bax, A., Kay, L., Driscoll, P., and Gronenborn, A. Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins. *Journal of the American Chemical Society*, 112(12):4989–4991, 1990.

Clore, G., Starich, M., and Gronenborn, A. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *Journal of the American Chemical Society*, 120(40):10571–10572, 1998.

Connely, G., Bai, Y., Jeng, M., and Englander, S. Isotope Effects in Peptide Group Hydrogen-Exchange. *Proteins-Structure Function and Genetics*, 17(1):87–92, 1993.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. WebLogo: A sequence logo generator. *Genome Research*, 14(6):1188–1190, 2004.

Dais, P., Tylianakis, E., Kanetakis, J., and Taravel, F. C-13 nuclear magnetic relaxation study of segmental dynamics of hyaluronan in aqueous solutions. *Biomacromolecules*, 6 (3):1397–1404, 2005.

Das, R. K. and Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Science of the United States of America*, 110(33):13392–13397, 2013.

De La Cruz, E., Ostap, E., Brundage, R., Reddy, K., Sweeney, H., and Safer, D. Thymosin-beta(4) changes the conformation and dynamics of actin monomers. *Biophysical Journal*, 78(5):2516–2527, 2000.

De Simone, A., Cavalli, A., Hsu, S.-T. D., Vranken, W., and Vendruscolo, M. Accurate Random Coil Chemical Shifts from an Analysis of Loop Regions in Native States of Proteins. *Journal of the American Chemical Society*, 131(45):16332+, 2009.

DeJean de la Batie, R., Laupretre, F., and Monnerie, L. Carbon-13 nmr investigation of local dynamics in bulk polymers at temperatures well above the glass transition temperature. 1. poly(vinyl methyl ether). *Macromolecules*, 21(7):2045–2052, 1988a.

DeJean de la Batie, R., Laupretre, F., and Monnerie, L. Carbon-13 nmr investigation of local dynamics in bulk polymers at temperatures well above the glass transition temperature. 2. poly(propylene oxide) and linear and cross-linked poly(ethylene oxides). *Macromolecules*, 21(7):2052–2058, 1988b.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. NMRPIPE - A Multidimensional Spectral Processing System Based on UNIX Pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 1995.

Derry, J., Ochs, H., and Francke, U. Isolation of a Novel Gene Mutated in Wiskott-Aldrich Syndrome. *Cell*, 78(4):635–644, 1994.

Didry, D., Cantrelle, F. X., Husson, C., Roblin, P., Moorthy, A. M. E., Perez, J., Clainche, C. L., Hertzog, M., Guittet, E., Carlier, M.-F., Heijenoort, C. V., and Renault, L. How a single residue in ß-thymosin and wh2 domains controls their functions in actin assembly. *EMBO Journal*, 31(4):1000–1013, 2012.

Disanza, A., Steffen, A., Hertzog, M., Frittoli, E., Rottner, K., and Scita, G. Actin polymerization machinery: the finish line of signaling networks, the starting point of cellular movement. *Cellular and Molecular Life Sciences*, 62(9):955–970, 2005.

Domanski, M., Hertzog, M., Coutant, J., Gutsche-Perelroizen, I., Bontems, F., Carlier, M., Guittet, E., and van Heijenoort, C. Coupling of folding and binding of thymosin beta 4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. *Journal of Biological Chemistry*, 279(22):23637–23645, 2004.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16):3433–3434, 2005.

Drugmand, J.-C., Schneider, Y.-J., and Agathos, S. N. Insect cells as factories for biomanufacturing. *Biotechnology Advances*, 30(5, SI):1140–1157, 2012.

Ducka, A. M., Joel, P., Popowicz, G. M., Trybus, K. M., Schleicher, M., Noegel, A. A., Huber, R., Holak, T. A., and Sitar, T. Structures of actin-bound Wiskott-Aldrich syndrome protein homology 2 (WH2) domains of Spire and the implication for filament nucleation. *Proceedings of the National Academy of Science of the United States of America*, 107 (26):11757–11762, 2010.

Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z., and Uversky, V. N. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*, 9(2), 2008.

Edwards, J. Are beta-thymosins WH2 domains? *FEBS Letters*, 573(1-3):231–232, 2004.

Egorova-Zachernyuk, T. A., Bosman, G. J. C. G. M., and DeGrip, W. J. Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. *Applied Microbiology and Biotechnology*, 89(2):397–406, 2011.

Elazari-Shalom, H., Shaked, H., Esteban-Martin, S., Salvatella, X., Barda-Saad, M., and Chill, J. H. New insights into the role of the disordered WIP N-terminal domain revealed by NMR structural characterization. *FEBS Journal*, 282(4):700–714, 2015.

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357(2):289–298, 2006.

Espinoza-Fonseca, M. L. Reconciling binding mechanisms of intrinsically disordered proteins. *Biochemical and Biophysical Research Communications*, 382(3):479–482, 2009.

Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. A., and Clore, G. M. Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature*, 480(7376):268–U161, 2011.

Fawzi, N. L., Ying, J., Torchia, D. A., and Clore, G. M. Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nature Protocols*, 7(8):1523–1533, 2012.

Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 51(4):219–242, 2007.

Fisher, E. *Berichte der Deutschen chemischen Gesellschaft zu Berlin*, 27:2985, 1894.

Fitzgerald, J. E., Jha, A. K., Sosnick, T. R., and Freed, K. F. Polypeptide motions are dominated by peptide group oscillations resulting from dihedral angle correlations between nearest neighbors. *Biochemistry*, 46(3):669–682, 2007.

Fizil, Á., Gáspári, Z., Barna, T., Marx, F., and Batta, G. "invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, cest-nmr experiments, and molecular dynamics calculations. *Chemistry – A European Journal*, 21 (13):5136–5144, 2015.

Forbes, J. G., Jin, A. J., Ma, K., Gutierrez-Cruz, G., Tsai, W. L., and Wang, K. Titin PEVK segment: charge-driven elasticity of the open and flexible polyampholyte. *Journal of Muscle Research and Cell Motility*, 26(6-8):291–301, 2005. International Symposium on Muscle Elastic Proteins, Chiba Univ, Chiba, JAPAN, NOV 19-21, 2004.

Frankel, S., Condeelis, J., and Leinwand, L. Expression of Action in Escherichia-coli - Aggregation, Solubilization, and Functional-Analysis. *Journal of Biological Chemistry*, 265(29):17980–17987, 1990.

Fujii, T., Iwane, A. H., Yanagida, T., and Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, 467(7316):724–U117, 2010.

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of Molecular Biology*, 338(5):1015–1026, 2004.

Fuxreiter, M., Toth-Petroczy, A., Kraut, D. A., Matouschek, A. T., Lim, R. Y. H., Xue, B., Kurgan, L., and Uversky, V. N. Disordered Proteinaceous Machines. *Chemical Reviews*, 114(13):6806–6843, 2014.

Galea, C. A., Wang, Y., Sivakolundu, S. G., and Kriwacki, R. W. Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*, 47(29):7598–7609, 2008.

Galkin, V. E., Orlova, A., Schroeder, G. F., and Egelman, E. H. Structural polymorphism in F-actin. *Nature Structural and Molecular Biology*, 17(11):1318–U169, 2010.

Galkin, V. E., Orlova, A., and Egelman, E. H. Actin Filaments as Tension Sensors. *Current Biology*, 22(3):R96–R101, 2012.

Galkin, V. E., Orlova, A., Vos, M. R., Schroeder, G. F., and Egelman, E. H. Near-Atomic Resolution for One State of F-Actin. *Structure*, 23(1):173–182, 2015.

Gallopin, M., Ochsenbein, F., Guittet, E., and van Heijenoort, C. Analysis of slow motions in the micro-millisecond range on domain 1 of annexin I. *Comptes Rendus Chimie*, 7(3-4): 253–258, 2004. 18th GERM Conference, Batz sur Mer, FRANCE, MAY 18-23, 2003.

Gallopin, M. *Analyse par RMN et Dichroïsme Circulaire de la dynamique, de la stabilité et du repliement du domaine 1 de l'annexine 1 humaine*. PhD thesis, Université Paris Sud, 2007.

Gans, P., Hamelin, O., Sounier, R., Ayala, I., Dura, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J., and Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angewandte Chemie-International Edition*, 49(11):1958–1962, 2010.

Gao, J., Li, Y., and Yan, H. NMR solution structure of domain 1 of human annexin I shows an autonomous folding unit. *Journal of Biological Chemistry*, 274(5):2971–2977, 1999.

Gardner, K. and Kay, L. Production and incorporation of N-15, C-13, H-2 (H-1-delta 1 methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society*, 119(32):7599–7600, 1997.

Gasca, S., Hill, D., J, K., and Rossant, J. Characterization of a Gene Trap Insertion Into a Novel Gene, Cordon-Bleu, Expressed in Axial Structures of the Gastrulating Mouse Embryo. *Developmental Genetics*, 17(2):141–154, 1995.

Gaucher, J.-F., Mauge, C., Didry, D., Guichard, B., Renault, L., and Carlier, M.-F. Interactions of Isolated C-terminal Fragments of Neural Wiskott-Aldrich Syndrome Protein (N-WASP) with Actin and Arp2/3 Complex. *Journal of Biological Chemistry*, 287(41): 34646–34659, 2012.

Gossert, A. D., Hinniger, A., Gutmann, S., Jahnke, W., Strauss, A., and Fernandez, C. A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *Journal of Biomolecular NMR*, 51(4):449–456, 2011.

Graceffa, P. and Dominguez, R. Crystal structure of monomeric actin in the ATP state - Structural basis of nucleotide-dependent actin dynamics. *Journal of Biological Chemistry*, 278(36):34172–34180, 2003.

Gsponer, J., Futschik, M. E., Teichmann, S. A., and Babu, M. M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science*, 322 (5906):1365–1368, 2008.

Guharoy, M., Szabo, B., Martos, S. C., Kosol, S., and Tompa, P. Intrinsic Structural Disorder in Cytoskeletal Proteins. *Cytoskeleton*, 70(10, SI):550–571, 2013.

Gutierrez-Cruz, G., Van Heerden, A., and Wang, K. Modular motif, structural folds and affinity profiles of the PEVK segment of human fetal skeletal muscle titin. *Journal of Biological Chemistry*, 276(10):7442–7449, 2001.

Guzman, L., Belin, D., Carson, M., and Beckwith, J. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-BAD Promoter. *Journal of Bacteriology*, 177(14):4121–4130, 1995.

Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. Introducing Protein Intrinsic Disorder. *Chemical Reviews*, 114(13):6561–6588, 2014.

Hall, C. and Helfand, E. Conformational State Relaxation in Polymers - Time-Correlation Functions. *Journal of Chemical Physics*, 77(6):3275–3282, 1982.

Hamatsu, J., O'Donovan, D., Tanaka, T., Shirai, T., Hourai, Y., Mikawa, T., Ikeya, T., Mishima, M., Boucher, W., Smith, B. O., Laue, E. D., Shirakawa, M., and Ito, Y. High-Resolution Heteronuclear Multidimensional NMR of Proteins in Living Insect Cells Using a Baculovirus Protein Expression System. *Journal of the American Chemical Society*, 135(5):1688–1691, 2013.

Hansen, A., Petros, A., Mazar, A., Pederson, T., Rueter, A., and Fesik, S. A Practical Method for Uniform Isotopic Labeling of Recombinant Proteins in Mammalian-Cells. *Biochemistry*, 31(51):12713–12718, 1992.

Hansen, D. F., Zhou, Z., Feng, H., Jenkins, L. M. M., Bai, Y., and Kay, L. E. Binding kinetics of histone chaperone chz1 and variant histone h2a.z-h2b by relaxation dispersion nmr spectroscopy. *Journal of Molecular Biology*, 387:1–9, 2009.

Hansen, D. F., Vallurupalli, P., and Kay, L. E. An improved (15)N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *Journal of Physical Chemistry B*, 112(19):5898–5904, 2008.

Hansen, J., Lu, X., Ross, E., and Woody, R. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *Journal of Biological Chemistry*, 281(4):1853–1856, 2006.

Hansen, M., Mueller, L., and Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*, 5(12): 1065–1074, 1998.

Heimsath, E. G., Jr. and Higgs, H. N. The C Terminus of Formin FMNL3 Accelerates Actin Polymerization and Contains a WH2 Domain-like Sequence That Binds Both Monomers and Filament Barbed Ends. *Journal of Biological Chemistry*, 287(5):3087–3098, 2012.

Helfand, E. Theory of the Kinetics of Conformational Transitions in Polymers. *Journal of Chemical Physics*, 54(11):4651–&, 1971.

Hemsath, L., Dvorsky, R., Fiegen, D., Carlier, M., and Ahmadian, M. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Molecular Cell*, 20(2):313–324, 2005.

Hendzel, M. J. The F-act's of nuclear actin. *Current Opinion in Cell Biology*, 28:84–89, 2014.

Hertzog, M., Yarmola, E., Didry, D., Bubb, M., and Carlier, M. Control of actin dynamics by proteins made of beta-thymosin repeats - The actobindin family. *Journal of Biological Chemistry*, 277(17):14786–14792, 2002.

Hertzog, M., van Heijenoort, C., Didry, D., Gaudier, M., Coutant, J., Gigant, B., Didelot, G., Preat, T., Knossow, M., Guittet, E., and Carlier, M. The beta-thymosin/WH2 domain: Structural basis for the switch from inhibition to promotion of actin assembly. *Cell*, 117 (5):611–623, 2004.

Holmes, K., Popp, D., Gebhard, W., and Kabsch, W. Atomic Model of the Actin Filament. *Nature*, 347(6288):44–49, 1990.

Holt, C. Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Current Opinion in Structural Biology*, 23(3):420–425, 2013.

Huff, T., Muller, C., Otto, A., Netzker, R., and Hannappel, E. beta-thymosins, small acidic peptides with multiple functions. *International Journal of Biochemistry and Cell Biology*, 33 (3):205–220, 2001.

Husson, C., Renault, L., Didry, D., Pantaloni, D., and Carlier, M.-F. Cordon-Bleu Uses WH2 Domains as Multifunctional Dynamizers of Actin Filament Assembly. *Molecular Cell*, 43(3):464–477, 2011.

Hwang, T., van Zijl, P., and Mori, S. Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of Biomolecular NMR*, 11(2):221–226, 1998.

Irobi, E., Aguda, A., Larsson, M., Guerin, C., Yin, H., Burtnick, L., Blanchoin, L., and Robinson, R. Structural basis of actin sequestration by thymosin-beta 4: implications for WH2 proteins. *EMBO Journal*, 23(18):3599–3608, 2004.

Ishii, Y., Markus, M., and Tycko, R. Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *Journal of Biomolecular NMR*, 21(2):141–151, 2001.

Iwasa, M., Maeda, K., Narita, A., Maeda, Y., and Oda, T. Dual roles of Gln(137) of actin revealed by recombinant human cardiac muscle alpha-actin mutants. *Journal of Biological Chemistry*, 283(30):21045–21053, 2008.

Jensen, M. R., Markwick, P. R. L., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P., and Blackledge, M. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using nmr dipolar couplings. *Structure*, 17(9):1169–1185, 2009.

Jensen, M. R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R. W. H., and Blackledge, M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Application to the molecular recognition element of Sendai virus nucleoprotein. *Journal of the American Chemical Society*, 130(25):8055–8061, 2008.

Jensen, M. R., Zweckstetter, M., Huang, J.-r., and Backledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chemical Reviews*, 114(13):6632–6660, 2014.

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., and Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *Assay and Drug Development Technologies*, 9(4):342–353, 2011.

Jiao, Y., Walker, M., Trinick, J., Pernier, J., Montaville, P., and Carlier, M.-F. Mutagenetic and Electron Microscopy Analysis of Actin Filament Severing by Cordon-Bleu, a WH2 Domain Protein. *Cytoskeleton*, 71(3):170–183, 2014.

Joel, P., Fagnant, P., and Trybus, K. Expression of a nonpolymerizable actin mutant in Sf9 cells. *Biochemistry*, 43(36):11554–11559, 2004.

Kabsch, W., Mannherz, H., Suck, D., Pai, E., and Holmes, K. Atomic-Structure of the Actin - DNase-I Complex. *Nature*, 347(6288):37–44, 1990.

Kang, H., Bradley, M. J., McCullough, B. R., Pierre, A., Grintsevich, E. E., Reisler, E., and De La Cruz, E. M. Identification of cation-binding sites on actin that drive polymerization and modulate bending stiffness. *Proceedings of the National Academy of Science of the United States of America*, 109(42):16923–16927, 2012.

Kang, H., Bradley, M. J., Elam, W. A., and De La Cruz, E. M. Regulation of Actin by Ion-Linked Equilibria. *Biophysical Journal*, 105(12):2621–2628, 2013.

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell*, 149(4):753–767, 2012.

Kelly, A., Kranitz, H., Dotsch, V., and Mullins, R. Actin binding to the central domain of WASP/Scar proteins plays a critical role in the activation of the arp2/3 complex. *Journal of Biological Chemistry*, 281(15):10589–10597, 2006.

Khaitlina, S. Functional specificity of actin isoforms. In *International Review of Cytology - A Survey of Cell Biology*, volume 202, pages 35–98. 2001.

Khanduja, N. and Kuhn, J. R. Processive acceleration of actin barbed-end assembly by N-WASP. *Molecular Biology of the Cell*, 25(1):55–65, 2014.

Kim, A., Kakalis, L., Abdul-Manan, M., Liu, G., and Rosen, M. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature*, 404(6774):151–158, 2000.

Kjaergaard, M. and Poulsen, F. M. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *Journal of Biomolecular NMR*, 50(2):157–165, 2011.

Kjaergaard, M., Brander, S., and Poulsen, F. M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of Biomolecular NMR*, 49(2): 139–149, 2011.

Koestler, S. A., Rottner, K., Lai, F., Block, J., Vinzenz, M., and Small, V. F- and G-Actin Concentrations in Lamellipodia of Moving Cells. *PLoS ONE*, 4(3), 2009.

Korn, E., Carlier, M., MF, and Pantaloni, D. Actin Polymerization and ATP Hydrolysis. *Science*, 238(4827):638–644, 1987.

Korzhnev, D., Billeter, M., Arseniev, A., and Orekhov, V. NMR studies of Brownian tumbling and internal motions in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 38(3):197–266, 2001.

Korzhnev, D. M., Vernon, R. M., Religa, T. L., Hansen, A. L., Baker, D., Fersht, A. R., and Kay, L. E. Nonnative Interactions in the FF Domain Folding Pathway from an Atomic Resolution Structure of a Sparsely Populated Intermediate: An NMR Relaxation Dispersion Study. *Journal of the American Chemical Society*, 133(28):10974–10982, 2011.

Koshland, D. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Science of the United States of America*, 44(2):98–104, 1958.

Kragelj, J., Ozenne, V., Blackledge, M., and Jensen, M. R. Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts. *ChemPhysChem*, 14(13, SI): 3034–3045, 2013.

Kragelj, J., Palencia, A., Nanao, M. H., Maurin, D., Bouvignies, G., Blackledge, M., and Jensen, M. R. Structure and dynamics of the MKK7-JNK signaling complex. *Proceedings of the National Academy of Science of the United States of America*, 112(11):3409–3414, 2015.

Lakomek, N.-A., Lange, O. F., Walter, K. F. A., Fares, C., Egger, D., Lunkenheimer, P., Meiler, J., Grubmueller, H., Becker, S., de Groot, B. L., and Griesinger, C. Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochemical Society Transactions*, 36:1433–1437, 2008.

Lappalainen, P. and Mattila, P. Reply to: Are beta-thymosins WH2 domains? *FEBS Letters*, 573(1-3):233, 2004.

Lee, S. H., Kerff, F., Chereau, D., Ferron, F., Klug, A., and Dominguez, R. Structural basis for the actin-binding function of missing-in-metastasis. *Structure*, 15(2):145–155, 2007.

LeMaster, D. and Kushlan, D. Dynamical mapping of E-coli thioredoxin via C-13 NMR relaxation analysis. *Journal of the American Chemical Society*, 118(39):9255–9264, 1996.

Letunic, I., Doerks, T., and Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*, 40(D1):D302–D305, 2012.

Levitt, M. Symmetrical Composite Pulse Sequences for NMR Population-Inversion .2. Compensation of resonance offset. *Journal of Magnetic Resonance*, 50(1):95–110, 1982.

Libich, D. S., Fawzi, N. L., Ying, J., and Clore, G. M. Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proceedings of the National Academy of Science*, 110(28):11361–11366, 2013.

Lim, J., Xiao, T., Fan, J., and Yang, D. An Off-Pathway Folding Intermediate of an Acyl Carrier Protein Domain Coexists with the Folded and Unfolded States under Native Conditions. *Angewandte Chemie-International Edition*, 53(9):2358–2361, 2014.

Lim, W. K., Rösgen, J., and Englander, S. W. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proceedings of the National Academy of Sciences*, 106(8):2595–2600, 2009.

Lipari, G. and Szabo, A. Model-Free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules .2. Analysis of experimentat results. *Journal of the American Chemical Society*, 104(17):4559–4570, 1982.

Lo, M., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical Biochemistry*, 332(1):153–159, 2004.

Long, D., Marshall, C. B., Bouvignies, G., Mazhab-Jafari, M. T., Smith, M. J., Ikura, M., and Kay, L. E. A Comparative CEST NMR Study of Slow Conformational Dynamics of Small GTPases Complexed with GTP and GTP Analogues. *Angewandte Chemie-International Edition*, 52(41):10771–10774, 2013.

Loria, J., Rance, M., and Palmer, A. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *Journal of the American Chemical Society*, 121(10):2331–2332, 1999.

Macleanfletcher, S. and Pollard, T. Identification of a Factor in Conventional Muscle Actin Preparations which Inhibits Actin Filament Self-Association. *Biochemical and Biophysical Research Communications*, 96(1):18–27, 1980.

Marley, J., Lu, M., and Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20(1):71–75, 2001.

Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Science*, 15(12):2795–2804, 2006.

Mattila, P., Salminen, M., Yamashiro, T., and Lappalainen, P. Mouse MIM, a tissue-specific regulator of cytoskeletal dynamics, interacts with ATP-actin monomers through its C-terminal WH2 domain. *Journal of Biological Chemistry*, 278(10):8452–8459, 2003.

McNulty, B., Young, G., and Pielak, G. Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *Journal of Molecular Biology*, 355(5):893–897, 2006.

Meola, A., Deville, C., Jeffers, S. A., Guardado-Calvo, P., Vasiliauskaite, I., Sizun, C., Girard-Blanc, C., Malosse, C., van Heijenoort, C., Chamot-Rooke, J., Krey, T., Guittet,

E., Petres, S., Rey, F. A., and Bontems, F. Robust and low cost uniform N-15-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *Journal of Structural Biology*, 188(1):71–78, 2014.

Meyer, B. and Peters, T. NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie-International Edition*, 42(8):864–890, 2003.

Miki, H., Miura, K., and Takenawa, T. N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskeletal rearrangement in a PIP2-dependent manner downstream of tyrosine kinases. *EMBO Journal*, 15(19):5326–5335, 1996.

Miller, B. M. and Trybus, K. M. Functional effects of nemaline myopathy mutations on human skeletal alpha-actin. *Journal of Biological Chemistry*, 283(28):19379–19388, 2008.

Millet, O., Loria, J., Kroenke, C., Pons, M., and Palmer, A. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *Journal of the American Chemical Society*, 122(12):2867–2877, 2000.

Modig, K., Juergensen, V. W., Lindorff-Larsen, K., Fieber, W., Bohr, H. G., and Poulsen, F. M. Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Letters*, 581(25):4965–4971, 2007.

Morimoto, D., Walinda, E., Fukada, H., Sou, Y.-S., Kageyama, S., Hoshino, M., Fujii, T., Tsuchiya, H., Saeki, Y., Arita, K., Ariyoshi, M., Tochio, H., Iwai, K., Namba, K., Komatsu, M., Tanaka, K., and Shirakawa, M. The unexpected role of polyubiquitin chains in the formation of fibrillar aggregates. *Nature Communications*, 6, 2015.

Mundia, M. M., Demers, R. W., Chow, M. L., Perieteanu, A. A., and Dawson, J. F. Subdomain Location of Mutations in Cardiac Actin Correlate with Type of Functional Change. *PLoS ONE*, 7(5), 2012.

Munoz, V. and Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nature Structural Biology*, 1(6):399–409, 1994.

Niesen, F. H., Berglund, H., and Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212–2221, 2007.

Noguchi, T. Q. P., Kanzaki, N., Ueno, H., Hirose, K., and Uyeda, T. Q. P. A novel system for expressing toxic actin mutants in Dictyostelium and purification and characterization of a dominant lethal yeast actin mutant. *Journal of Biological Chemistry*, 282(38):27721–27727, 2007.

Nurnberg, A., Kitzing, T., and Grosse, R. Nucleating actin for invasion. *Nature Reviews Cancer*, 11(3):177–187, 2011.

Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., Pan, A. C., Liu, C. W., Fung, J. J., Bokoch, M. P., Thian, F. S., Kobilka, T. S., Shaw, D. E., Mueller, L., Prosser, R. S., and Kobilka, B. K. The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell*, 152(3):532–542, 2013.

Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztanyi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., and Gough, J. (DP2)-P-2: database of disordered protein predictions. *Nucleic Acids Research*, 41(D1):D508–D516, 2013.

Ochsenbein, F., Neumann, J. M., Guittet, E., and van Heijenoort, C. Dynamical characterization of residual and non-native structures in a partially folded protein by 15n nmr relaxation using a model based on a distribution of correlation times. *Protein Science*, 11:957–964, 2002.

Oda, T., Iwasa, M., Aihara, T., Maeda, Y., and Narita, A. The nature of the globular-to fibrous-actin transition. *Nature*, 457(7228):441–445, 2009.

Ogino, S., Kubo, S., Umemoto, R., Huang, S., Nishida, N., and Shimada, I. Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-Forming Toxin, Streptolysin O. *Journal of the American Chemical Society*, 131(31):10834+, 2009.

Ohki, T., Mikhailenko, S. V., Arai, T., Ishii, S., and Ishiwata, S. Improvement of the yields of recombinant actin and myosin V-HMM in the insect cell/baculovirus system by the addition of nutrients to the high-density cell culture. *Journal of Muscle Research and Cell Motility*, 33(5):351–358, 2012.

Ohman, L., Ljunggren, J., and Haggstrom, L. Induction of a Metabolic Switch in Insect Cells by Substrate Limited Fed-Batch Cultures. *Applied Microbiology and Biotechnology*, 43(6):1006–1013, 1995.

Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*, 9(1), 2008.

Oldfield, C., Cheng, Y., Cortese, M., Romero, P., Uversky, V., and Dunker, A. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, 44(37):12454–12470, 2005.

Orekhov, V. Y. and Jaravine, V. A. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 59(3): 271–292, 2011.

Otterbein, L., Graceffa, P., and Dominguez, R. The crystal structure of uncomplexed actin in the ADP state. *Science*, 293(5530):708–711, 2001.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-r., Jensen, M. R., Segard, S., Bernado, P., Charavay, C., and Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, 28(11):1463–1470, 2012a.

Ozenne, V., Schneider, R., Yao, M., Huang, J.-r., Salmon, L., Zweckstetter, M., Jensen, M. R., and Blackledge, M. Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *Journal of the American Chemical Society*, 134(36): 15138–15148, 2012b.

Padrick, S. B. and Rosen, M. K. Physical Mechanisms of Signal Integration by WASP Family Proteins. In Kornberg, RD and Raetz, CRH and Rothman, JE and Thorner, JW, editor, *Annual Review of Biochemistry, VOL 79*, volume 79 of *Annual Review of Biochemistry*, pages 707–735. Annual Reviews, 2010.

Panchal, S., Kaiser, D., Torres, E., Pollard, T., and Rosen, M. A conserved amphipathic helix in WASP/Scar proteins is essential for activation of Arp2/3 complex. *Nature Structural Biology*, 10(8):591–598, 2003.

Papayannopoulos, V., Co, C., Prehoda, K., Snapper, S., Taunton, J., and Lim, W. A polybasic motif allows N-WASP to act as a sensor of PIP2 density. *Molecular Cell*, 17(2):181–191, 2005.

Pardee, J. and Spudich, J. Mechanism of K+-Induced Actin Assembly. *Journal of Cell Biology*, 93(3):648–654, 1982.

Paunola, E., Mattila, P., and Lappalainen, P. WH2 domain: a small, versatile adapter for actin monomers. *FEBS Letters*, 513(1, SI):92–97, 2002.

Pernier, J., Orban, J., Avvaru, B. S., Jegou, A., Romet-Lemonne, G., Guichard, B., and Carlier, M.-F. Dimeric WH2 domains in Vibrio VopF promote actin filament barbed-end uncapping and assisted elongation. *Nature Structural and Molecular Biology*, 20(9):1069+, 2013.

Phan, I., Boyd, J., and Campbell, I. Dynamic studies of a fibronectin type I module pair at three frequencies: Anisotropic modelling and direct determination of conformational exchange. *Journal of Biomolecular NMR*, 8(4):369–378, 1996.

Pinder, J., Sleep, J., Bennett, P., and Gratzer, W. Concentrated Tris Solutions for the Preparation, Depolymerization, and Assay of Actin - Application to Erythroid Actin. *Analytical Biochemistry*, 225(2):291–295, 1995.

Pivovarova, A. V., Chebotareva, N. A., Kremneva, E. V., Lappalainen, P., and Levitsky, D. I. Effects of Actin-Binding Proteins on the Thermal Stability of Monomeric Actin. *Biochemistry*, 52(1):152–160, 2013.

Placial, J.-P. *Étude d'états multiples dans les protéines par Résonance Magnétique Nucléaire. Méthodologie du traitement des données de dispersion de relaxation.* PhD thesis, Université Pierre et Marie Curie, 2013.

Pollard, T. and Borisy, G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell*, 112(4):453–465, 2003.

Pollard, T., Blanchoin, L., and Mullins, R. Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annual Review of Biophysics and Biomolecular Structure*, 29: 545–576, 2000.

Pollard, T. D. and Cooper, J. A. Actin, a Central Player in Cell Shape and Movement. *Science*, 326(5957):1208–1212, 2009.

Prestegard, J. H., Sahu, S. C., Nkari, W. K., Morris, L. C., Live, D., and Gruta, C. Chemical shift prediction for denatured proteins. *Journal of Biomolecular NMR*, 55(2):201–209, 2013.

Rasia, R., Lescop, E., Palatnik, J., Boisbouvier, J., and Brutscher, B. Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *Journal of Biomolecular NMR*, 51(3):369–378, 2011.

Rebowski, G., Boczkowska, M., Hayes, D. B., Guo, L., Irving, T. C., and Dominguez, R. X-ray scattering study of actin polymerization nuclei assembled by tandem W domains. *Proceedings of the National Academy of Science of the United States of America*, 105(31):10785–10790, 2008.

Rebowski, G., Namgoong, S., Boczkowska, M., Leavis, P. C., Navaza, J., and Dominguez, R. Structure of a Longitudinal Actin Dimer Assembled by Tandem W Domains: Implications for Actin Filament Nucleation. *Journal of Molecular Biology*, 403(1):11–23, 2010.

Renault, L., Deville, C., and van Heijenoort, C. Structural Features and Interfacial Properties of WH2, beta-Thymosin Domains and Other Intrinsically Disordered Domains in the Regulation of Actin Cytoskeleton Dynamics. *Cytoskeleton*, 70(11, SI):686–705, 2013.

Romero, P., Obradovic, Z., and Dunker, A. Natively disordered proteins: functions and predictions. *Applied Bioinformatics*, 3:105–113, 2004.

Rouayrenc, J. and Travers, F. The 1st Step in the Polymerization of Actin. *European Journal of Biochemistry*, 116(1):73–77, 1981.

Rould, M. A., Wan, Q., Joel, P. B., Lowey, S., and Trybus, K. M. Crystal structures of expressed non-polymerizable monomeric actin in the ADP and ATP states. *Journal of Biological Chemistry*, 281(42):31909–31919, 2006.

Rozsa, G., Szentgyorgyi, A., and Wyckoff, R. The Electron Microscopy of F-Actin. *Biochimica et Biophysica Acta*, 3(5-6):561–569, 1949.

Ruckert, M. and Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*, 122(32):7793–7797, 2000.

Rutkevich, L., Teal, D., and Dawson, J. Expression of actin mutants to study their roles in cardiomyopathy. *Canadian Journal of Physiology and Pharmacology*, 84(1):111–119, 2006. National Research Forum for Young Investigators in Circulatory and Respiratory Health, Winnipeg, CANADA, APR 28-MAY 01, 2005.

Safer, D., Sosnick, T., and Elzinga, M. Thymosin beta(4) binds actin in an extended conformation and contacts both the barbed and pointed ends. *Biochemistry*, 36(19):5806–5816, 1997.

Sastry, M., Xu, L., Georgiev, I. S., Bewley, C. A., Nabel, G. J., and Kwong, P. D. Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy. *Journal of Biomolecular NMR*, 50(3):197–207, 2011.

Schafer, D., Jennings, P., and Cooper, J. Rapid and efficient purification of actin from nonmuscle sources. *Cell Motility and the Cytoskeleton*, 39(2):166–171, 1998.

Schanda, P., Van Melckebeke, H., and Brutscher, B. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society*, 128(28): 9042–9043, 2006.

Schneider, R., Maurin, D., Communie, G., Kragelj, J., Hansen, D. F., Ruigrok, R. W. H., Jensen, M. R., and Blackledge, M. Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *Journal of the American Chemical Society*, 137(3):1220–1229, 2015.

Schueler, S., Hauptmann, J., Perner, B., Kessels, M. M., Englert, C., and Qualmann, B. Ciliated sensory hair cell formation and function require the F-BAR protein syndapin I and the WH2 domain-based actin nucleator Cobl. *Journal of Cell Science*, 126(1):196–208, 2013.

Schuler, H., Lindberg, U., Schutt, C., and Karlsson, R. Thermal unfolding of G-actin monitored with the DNase I-inhibition assay - Stabilities of actin isoforms. *European Journal of Biochemistry*, 267(2):476–486, 2000.

Schwalbe, H., Fiebig, K., Buck, M., Jones, J., Grimshaw, S., Spencer, A., Glaser, S., Smith, L., and Dobson, C. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, 36(29):8977–8991, 1997.

Schwarzinger, S., Kroon, G., Foss, T., Wright, P., and Dyson, H. Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. *Journal of Biomolecular NMR*, 18(1):43–48, 2000.

Schwarzinger, S., Kroon, G., Foss, T., Chung, J., Wright, P., and Dyson, H. Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society*, 123(13):2970–2978, 2001.

Seidel, S. A. I., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D., and Duhr, S. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, 59(3):301–315, 2013.

Selden, L., Kinosian, H., Estes, J., and Gershman, L. Impact of profilin on actin-bound nucleotide exchange and actin polymerization dynamics. *Biochemistry*, 38(9):2769–2778, 1999.

Sept, D. and McCammon, J. Thermodynamics and kinetics of actin filament nucleation. *Biophysical Journal*, 81(2):667–674, 2001.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223, 2009.

Sheterline, P., Clayton, J., and Sparrow, J. Actin. *Protein Profile*, 2(1):1–&, 1995.

Shoemaker, B., Portman, J., and Wolynes, P. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Science of the United States of America*, 97(16):8868+, 2000.

Sitar, T., Gallinger, J., Ducka, A. M., Ikonen, T. P., Wohlhoefler, M., Schmoller, K. M., Bausch, A. R., Joel, P., Trybus, K. M., Noegel, A. A., Schleicher, M., Huber, R., and Holak, T. A. Molecular architecture of the Spire-actin nucleus and its implication for actin filament assembly. *Proceedings of the National Academy of Science of the United States of America*, 108(49):19575–19580, 2011.

Skrynnikov, N., Dahlquist, F., and Kay, L. Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *Journal of the American Chemical Society*, 124(41):12352–12360, 2002.

Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., and Schuler, B. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proceedings of the National Academy of Science of the United States of America*, 111(13):4874–4879, 2014.

Spudich, J. and Watt, S. Regulation of Rabbit Skeletal Muscle Contraction .1. Biochemical Studies of Interaction of Tropomyosin-Troponin Complex with Actin and Proteolytic Fragments of Myosin. *Journal of Biological Chemistry*, 246(15):4866–&, 1971.

Sribenja, S., Wongkham, S., Wongkham, C., Yao, Q., and Chen, C. Roles and Mechanisms of beta-Thymosins in Cell Migration and Cancer Metastasis: An Update. *Cancer Investigation*, 31(2):103–110, 2013.

Stoll, R., Voelter, W., and Holak, T. Conformation of thymosin beta(9) in water/fluoroalcohol solution determined by NMR spectroscopy. *Biopolymers*, 41(6):623–634, 1997.

Stolp, B. and Fackler, O. T. How HIV Takes Advantage of the Cytoskeleton in Entry and Replication. *Viruses-Basel*, 3(4):293–311, 2011.

Straub, F. Actin. *Studies from the Institute of Medical Chemistry University Szeged*, 2:3–15, 1942.

Strauss, A., Bitsch, F., Fendrich, G., Graff, P., Knecht, R., Meyhack, B., and Jahnke, W. Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *Journal of Biomolecular NMR*, 31(4):343–349, 2005.

Sugase, K., Dyson, H. J., and Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–U11, 2007.

Tamiola, K., Acar, B., and Mulder, F. A. A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society*, 132 (51):18000–18003, 2010.

Tamura, M., Ito, K., Kunihiro, S., Yamasaki, C., and Haragauchi, M. Production of human beta-actin and a mutant using a bacterial expression system with a cold shock vector. *Protein Expression and Purification*, 78(1):1–5, 2011.

Theillet, F.-X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., Li, C., Crowley, P. B., Gierasch, L., Pielak, G. J., Elcock, A. H., Gershenson, A., and Selenko, P. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chemical Reviews*, 114(13):6661–6714, 2014.

Tilney, L., Derosier, D., and Tilney, M. How Listeria Exploits Host-Cell Actin to Form its own Cytoskeleton .1. Formation of a Tail and How that Tail Might be Involved in Movement. *Journal of Cell Biology*, 118(1):71–81, 1992a.

Tilney, L., Derosier, D., Weber, A., and Tilney, M. How Listeria Exploits Host-Cell Actin to Form its own Cytoskeleton .2. Nucleation, Actin Filament Polarity, Filament Assembly, and Evidence for a Pointed End Capper. *Journal of Cell Biology*, 118(1):83–93, 1992b.

Tjandra, N. and Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 278(5340):1111–1114, 1997.

Tolkatchev, D., Xu, P., and Ni, F. Probing the kinetic landscape of transient peptide-protein interactions by use of peptide 15n nmr relaxation dispersion spectroscopy: Binding of

an antithrombin peptide to human prothrombin. *Journal of the American Chemical Society*, 125:12432–12442, 2003.

Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25(9): 847–855, 2003.

Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579(15):3346–3354, 2005.

Tompa, P. and Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, 33(1):2–8, 2008.

Uversky, V. N. and Dunker, A. K. *Intrinsically Disordered Protein Analysis*, volume 895 of *Methods in Molecular Biology*. Humana Press, 2012.

Uversky, V., Oldfield, C., and Dunker, A. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*, 18(5):343–384, 2005.

Vajpai, N., Strauss, A., Fendrich, G., Cowan-Jacob, S. W., Manley, P. W., Grzesiek, S., and Jahnke, W. Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *Journal of Biological Chemistry*, 283(26):18292–18302, 2008.

Vallurupalli, P., Scott, L., Williamson, J. R., and Kay, L. E. Strong coupling effects during X-pulse CPMG experiments recorded on heteronuclear ABX spin systems: artifacts and a simple solution. *Journal of Biomolecular NMR*, 38(1):41–46, 2007.

Vallurupalli, P., Bouvignies, G., and Kay, L. E. Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *Journal of the American Chemical Society*, 134(19):8148–8161, 2012.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews*, 114(13):6589–6631, 2014.

Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A. K., Felli, I. C., Forman-Kay, J. D., Kriwacki, R. W., Pierattelli, R., Sussman, J., Svergun, D. I., Uversky, V. N., Vendruscolo, M., Wishart, D., Wright, P. E., and Tompa, P. pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Research*, 42(D1):D326–D335, 2014.

Volkman, L., Storm, K., Aivazachvili, V., and Oppenheimer, D. Overexpression of actin in AcMNPV-infected cells interferes with polyhedrin synthesis and polyhedra formation. *Virology*, 225(2):369–376, 1996.

Vranken, W., Boucher, W., Stevens, T., Fogh, R., Pajon, A., Llinas, P., Ulrich, E., Markley, J., Ionides, J., and Laue, E. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics*, 59(4):687–696, 2005.

Walton, W. J., Kasprzakb, A. J., Hare, J. T., and Logan, T. M. An economic approach to isotopic enrichment of glycoproteins expressed from Sf9 insect cells. *Journal of Biomolecular NMR*, 36(4):225–233, 2006.

Wang, H., Robinson, R. C., and Burtnick, L. D. The Structure of Native G-Actin. *Cytoskeleton*, 67(7):456–465, 2010.

Wang, Y. and Jardetzky, O. Investigation of the neighboring residue effects on protein chemical shifts. *Journal of the American Chemical Society*, 124(47):14075–14084, 2002a.

Wang, Y. and Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Science*, 11(4):852–861, 2002b.

Ward, J., Sodhi, J., McGuffin, L., Buxton, B., and Jones, D. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, 337(3):635–645, 2004.

Wayt, J. and Bretscher, A. Cordon Bleu serves as a platform at the basal region of microvilli, where it regulates microvillar length through its WH2 domains. *Molecular Biology of the Cell*, 25(18):2817–2827, 2014.

Weber, S. C. and Brangwynne, C. P. Getting RNA and Protein in Phase. *Cell*, 149(6): 1188–1191, 2012.

Wishart, D. and Sykes, B. Chemical-Shifts as a Tool for Structure Determination. In *Nuclear Magnetic Resonance, PT C*, volume 239 of *Methods in Enzymology*, pages 363–392. 1994.

Wishart, D., Bigam, C., Holm, A., Hodges, R., and Sykes, B. H-1, C-13 AND N-15 Random Coil NMR Chemical-Shifts of the Common Amino-Acids .1. Investigations of Nearest-Neighbor Effects. *Journal of Biomolecular NMR*, 5(1):67–81, 1995.

Wright, P. and Dyson, H. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2):321–331, 1999.

Wright, P. E. and Dyson, H. J. Linking folding and binding. *Current Opinion in Structural Biology*, 19(1):31–38, 2009.

Wright, P. E. and Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29, 2015.

Xue, B. and Robinson, R. C. Guardians of the actin monomer. *European Journal of Cell Biology*, 92(10–11):316 – 332, 2013. {SPP} 1464:Actin Dynamics – Part I.

Xue, B., Leyrat, C., Grimes, J. M., and Robinson, R. C. Structural basis of thymosin-beta 4/profilin exchange leading to actin filament polymerization. *Proceedings of the National Academy of Science of the United States of America*, 111(43):E4596–E4605, 2014.

Yamaguchi, H., Miki, H., Suetsugu, S., Ma, L., Kirschner, M., and Takenawa, T. Two tandem verprolin homology domains are necessary for a strong activation of Arp2/3 complex-induced actin polymerization and induction of microspike formation by N-WASP. *Proceedings of the National Academy of Science of the United States of America*, 97 (23):12631–12636, 2000.

Yates, S. P., Otley, M. D., and Dawson, J. F. Overexpression of cardiac actin with baculovirus is promoter dependent. *Archives of Biochemistry and Biophysics*, 466(1):58–65, 2007.

Zhang, H., Neal, S., and Wishart, D. RefDB: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR*, 25(3):173–195, 2003.

Zhou, J. and van Zijl, P. C. M. Chemical exchange saturation transfer imaging and spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48(2-3):109–136, 2006.

IV

Annexes

Plasmid maps

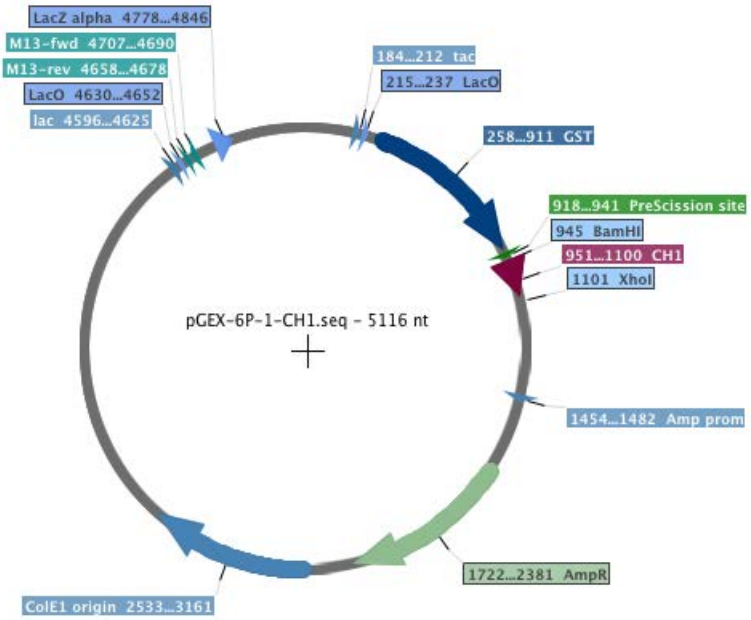


Figure 13 – *pGEX-6P-1-(CH1/CH2) for overexpression of CH1 and CH2 in fusion with a N-terminal GST+Prescision cleavage site*

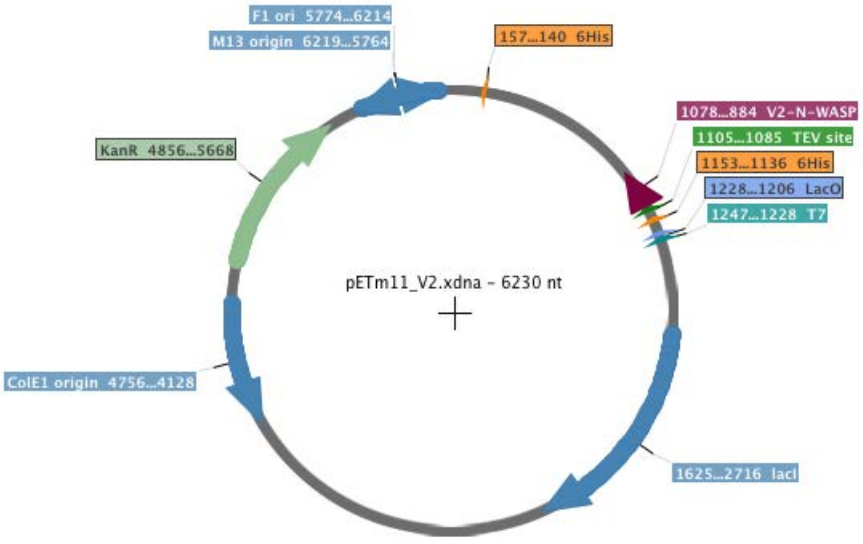


Figure 14 – *pETm11-(NWASPV2) for overexpression of NWASPV2 in fusion with a N-terminal 6His tag+TEV cleavage site*

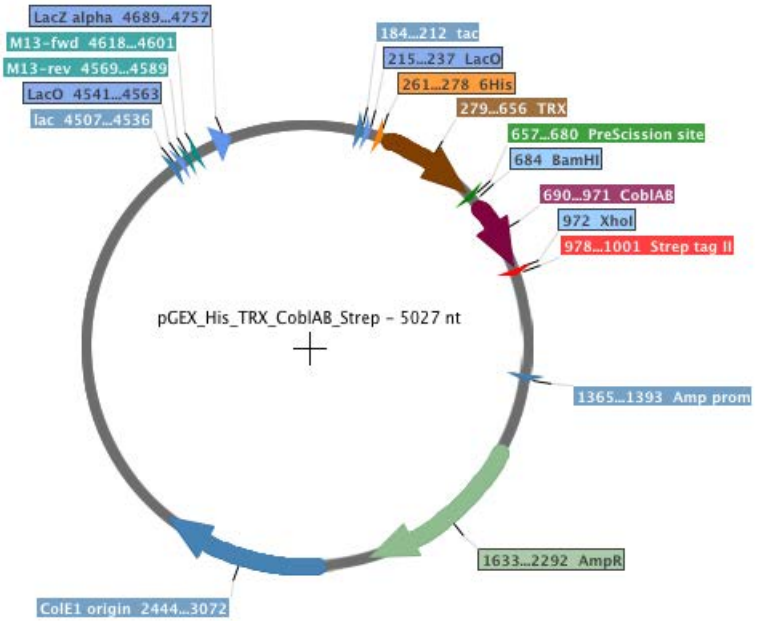


Figure 15 – *pGEXmod-His-TRX-(CoblAB)-Strep for overexpression of CoblAB in fusion with a N-terminal 6His tag+TRX+Prescision cleavage site and with a C-terminal Streptag*

Buffers composition

Culture media

M9 medium

Na_2HPO_4	45 mM
KH_2PO_4	22 mM
NaCl	8.5 mM
NH_4Cl	18.5 mM
$MgSO_4$	2 mM
$CaCl_2$	10 μM
Glucose	13.9-22.2 mM
Thiamine	16.6 μM

$^{15}NH_4Cl$ and ^{13}C-glucose are used for ^{15}N- and ^{13}C-labeling respectively. For deuteration, H_2O is replaced by D_2O.

LB medium

Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L

Purification buffers

CH1/CH2					
Lysis buffer		**Cleavage buffer**		**Gel-filtration buffer**	
Tris-HCl	50 mM	Tris-HCl	50 mM	Tris-HCl	50 mM
NaCl	0.5 M	NaCl	0.15 M	$CaCl_2$	0.5mM
$MgCl_2$	1 mM	$MgCl_2$	50 mM	DTT	1 mM
DTT	1 mM	DTT	1 mM		pH = 7
antiprotease		EDTA	0.5mM		
	pH = 7.5		pH = 7		

NWASPV2					
Lysis buffer		**Histrap buffer A**		**Gel-filtration buffer**	
Tris-HCl	100mM	Tris-HCl	50 mM	Tris-HCl	50 mM
NaCl	1 M	NaCl	1 M	$CaCl_2$	0.5mM
Imidazole	15 mM	Imidazole	15 mM	DTT	1 mM
EDTA	0.5 mM	DTT	1 mM		pH = 7
DTT	1 mM		pH = 8		
antiprotease		**Histrap buffer B**			
	pH = 8.2	500 mM	Imidazole		

CoblAB					
Lysis buffer		**Histrap buffer A**		**Cleavage buffer**	
PBS	1X	Tris-HCl	50 mM	Tris-HCl	50 mM
NaCl	1 M	NaCl	1 M	NaCl	0.15 M
Imidazole	15 mM	Imidazole	15 mM	Sucrose	1%
EDTA	0.5 mM	Sucrose	1%		pH = 7.5
Urea	0.5 M	DTT	1 mM	**Gel-filtration buffer**	
Sucrose	1%		pH = 7.4	Tris-HCl	50 mM
DTT	1 mM	**Histrap buffer B**		$CaCl_2$	0.5mM
antiprotease		500 mM	Imidazole	DTT	1 mM
	pH = 7.4				pH = 7

Actin sequence alignment

Figure 16 – *Sequence alignment of drosophila 5C actin (Dm-ACT1) human γ-cytoplasmic actin (Hs-ACTG) and human α-skeletal actin.*

Publications

- Renault, L., Deville, C., and van Heijenoort, C. Structural Features and Interfacial Properties of WH2, beta-Thymosin Domains and Other Intrinsically Disordered Domains in the Regulation of Actin Cytoskeleton Dynamics. *Cytoskeleton*, 70(11, SI):686–705, 2013.
- Meola, A., Deville, C., Jeffers, S. A., Guardado-Calvo, P., Vasiliauskaite, I., Sizun, C., Girard-Blanc, C., Malosse, C., van Heijenoort, C., Chamot-Rooke, J., Krey, T., Guittet, E., Petres, S., Rey, F. A., and Bontems, F. Robust and low cost uniform N-15-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *Journal of Structural Biology*, 188(1):71–78, 2014.
- Deville, C., Girard-Blanc, C., Assrir, N., Renault, L. Bontems, F., Petres, S. and van Heijenoort, C. Mutations in a non-polymerizable actin mutant partially disrupt interactions with β-thymosin domains. *(to be submitted)*

REVIEW ARTICLE
Cytoskeleton, November 2013 70:686–705 (doi: 10.1002/cm.21140)

Structural Features and Interfacial Properties of WH2, β-Thymosin Domains and Other Intrinsically Disordered Domains in the Regulation of Actin Cytoskeleton Dynamics

Louis Renault,[1]* Célia Deville,[2] and Carine van Heijenoort[2]

[1]*Laboratoire d'Enzymologie et Biochimie Structurales, Centre de Recherche de Gif, CNRS, Gif-sur-Yvette, France*
[2]*Laboratoire de Chimie et Biologie Structurales, Institut de Chimie des Substances Naturelles, Centre de Recherche de Gif, CNRS, Gif-sur-Yvette, France*

Received 10 June 2013; Revised 28 August 2013; Accepted 1 September 2013
Monitoring Editor: Miklós Nyitrai

Many actin-binding proteins (ABPs) use complex multidomain architectures to integrate and coordinate multiple signals and interactions with the dynamic remodeling of actin cytoskeleton. In these proteins, small segments that are intrinsically disordered in their unbound native state can be functionally as important as identifiable folded units. These functional intrinsically disordered regions (IDRs) are however difficult to identify and characterize *in vitro*. Here, we try to summarize the state of the art in understanding the structural features and interfacial properties of IDRs involved in actin self-assembly dynamics. Recent structural and functional insights into the regulation of widespread, multifunctional WH2/β-thymosin domains, and of other IDRs such as those associated with WASP/WAVE, formin or capping proteins are examined. Understanding the functional versatility of IDRs in actin assembly requires apprehending by multiple structural and functional approaches their large conformational plasticity and dynamics in their interactions. In many modular ABPs, IDRs relay labile interactions with multiple partners and act as interaction hubs in interdomain and protein–protein interfaces. They thus control multiple conformational transitions between the inactive and active states or between various active states of multidomain ABPs, and play an important role to coordinate the high turnover of interactions in actin self-assembly dynamics. © 2013 Wiley Periodicals, Inc.

Key Words: actin cytoskeleton remodeling; actin-binding proteins; WH2 (Wiskott-Aldrich syndrome protein Homology domain 2) domains; ß-thymosin domains; intrinsically disordered proteins; functional versatility; multifunctionality; structural biology

*Address Correspondence to: Louis Renault; Laboratoire d'Enzymologie et Biochimie Structurales, Centre de Recherche de Gif, CNRS, Gif-sur-Yvette, France. E-mail: renault@lebs.cnrs-gif.fr
Published online 22 October 2013 in Wiley Online Library (wileyonlinelibrary.com).

Introduction

The actin cytoskeleton forms a very dynamic structural network that is constantly remodelled in eukaryotic cells to control and coordinate multiple cellular processes, including cell polarity establishment and maintenance, polarized cell migration, cell adhesion, cytokinesis, intracellular transport, or intracellular pathogen infections. These cellular processes are driven by the tightly regulated self-assembly of actin monomers (G-actin) in polarized, double-stranded helical filaments (F-actin), formation of higher order F-actin structures, and their controlled disassembly [Winder and Ayscough, 2005; Pollard and Cooper, 2009]. The dynamic behavior of the actin cytoskeleton and its involvement in many diverse cellular processes rely on a large array of actin-binding proteins (ABPs) that work in concert. Each ABP regulates elementary functions with G- or F-actin. ABPs regulate for example the nucleotide-binding and self-assembly properties of the cellular G-actin pool or actin filament nucleation, polarized elongation, length, stability, disassembly, cross-linking or anchoring. These multiple activities are spatiotemporally regulated within the cells downstream of specific cellular cues to control actin dynamics at restricted cellular locations. To adapt cell migration to diverse external microenvironments, the cortical actin cytoskeleton is for example intimately connected with the plasma membrane and signaling emanating from plasma membrane receptors, curvature, composition or dynamics [Bisi et al., 2013]. Changes in the actin cytoskeleton organization are also controlled by changes in the transcriptional expression profile of ABPs. Finally, post-translational regulations of ABPs or their partners bring additional labile informations in proteins that are thought to drive rapid changes in actin dynamics [Campellone and

Welch, 2010; Hannappel, 2010; Hansen and Kwiatkowski, 2013; Mendoza, 2013]. Many ABPs are therefore made of multiple functional folded domains by which they integrate and coordinate multiple cellular signals and interactions. Understanding all their interactions and coordination at the molecular level is arduous and requires to combine many functional and structural investigations. In the last years, an additional level of complexity has been highlighted in many studies on multidomain ABP regulation: beyond identifiable folded domains, modular ABPs can contain small regions or segments that are intrinsically disordered in their native state but essential for maintaining the ensemble of interactions and activities of the full-length protein. These functional intrinsically disordered segments/regions are difficult to identify and characterize *in vitro* because they frequently have small sizes, high sequence variability and display inherent instability, plasticity and dynamics in solution. In addition they are especially sensitive to proteolysis and conducive to protein aggregations. Deciphering their specific structural properties and mechanisms of regulations in modular ABPs represents nowadays a challenging but important task to understand the full regulation of many multidomain protein architectures involved in actin cytoskeleton remodeling. We overview here the hallmarks of several cytoskeletal intrinsically disordered proteins (IDPs) or segments/regions (IDRs) involved in various important functions of actin self-assembly dynamics.

About 10–35% of prokaryotic and 15–45% of eukaryotic proteins contain significant disorder, that is, long disordered regions of at least 30 amino acids [Uversky and Dunker, 2010; Tompa, 2012]. Their higher presence in eukaryotes is correlated with their frequent involvement in signal transduction or transcriptional regulation. Many further investigations are required to understand this new class of proteins that defy the structure-function paradigm. The distinguishing feature of functional IDPs/IDRs is their inability to fold into a unique and stable tertiary structure in solution and/or in the crowding environment of cells [Tompa, 2011]. Their native isolated state consists in a dynamic ensemble of interconverting conformers. Although some adopt well defined secondary or tertiary structures upon binding to their partners, numerous IDPs appear to be functional in complexes through retaining various degree of dynamics and disorder, or adopting multiple conformations in protein–protein interfaces, leading to the introduction of fuzzy complexes [Fuxreiter, 2012; Fuxreiter and Tompa, 2012]. The functional features of IDPs can be further summarized by adaptability in binding, often weak but specific binding, and frequent regulation by post-translational modification or alternative splicing. The interest to understand deeply the molecular mechanisms of IDP/IDR regulations at protein–protein interfaces is further accentuated by the involvement of numerous IDPs in many human diseases, such as amyloidoses and neurodegenerative diseases, but also cancer, cardiovascular disease, and diabetes [Uversky, 2010]. What are the main functions, interfacial properties, and regulatory mechanisms of many IDPs/IDRs involved in cytoskeleton remodeling, and their commonalities and uniqueness compared with other IDPs remain open issues.

In this journal issue, Guharoy et al. provide a thorough overview on the abundance, functional diversity, and implications in diseases of IDPs/IDRs involved in the regulation of the three major cytoskeletal systems of eukaryotic cells [Guharoy et al., 2013]. Here, we try to capture the state of the art in understanding the main structural features and interfacial properties of IDRs involved in actin self-assembly dynamics. We outline more especially recent molecular findings that analyze the regulation of WH2 (WASP-homology 2) and β-thymosin (βT) domains in actin assembly, because they represent widespread, archetypal actin-binding IDRs [Dominguez, 2009; Carlier et al., 2011]. Recent findings on IDRs that are associated with the regulations of nucleation promoting factors (NPFs) of the WASP/Scar/WAVE family, formin, or capping proteins (CPs) are also examined. We review what the main structural and functional properties of these IDRs are when they work in actin assembly as single domains, alone or in association with other domains targeting actin or other ABPs, or as multiple repeats. Our analysis highlights how IDPs/IDRs involved in actin cytoskeleton remodeling are important and adapted for the multiple and rapid changes in interdomain interfaces and protein–protein interactions that control the high turnover of interactions in actin assembly dynamics.

Elementary Properties and Functional Versatility of Isolated Single Actin Cytoskeletal IDRs: Example of WH2/βT Domains

Among actin cytoskeletal IDRs/IDPs, widespread WH2 and βT domains are among the shortest functional actin-binding modules known. They either build mono-domain IDPs or exist as single or repeated units in many multidomain proteins.

Variability of WH2/ßT IDRs and of Their Structural Organizations in Proteins

Individual ß-thymosin domains were first identified as a family of ~5 kDa, heat stable polypeptide isoforms from calf thymus in the mid-1960's, called ß-thymosins [Huff et al., 2001]. Within cells, ß-thymosins mainly interact with actin monomers and influence many cellular functions, including cell migration or endothelial cell attachment and spreading [Sribenja et al., 2013]. Thymosin-ß4 (Tß4), the most abundant ß-thymosin isoform in metazoans, is present at very high concentrations (300–500 μM) in platelets and lymphocytes, and is overexpressed in several

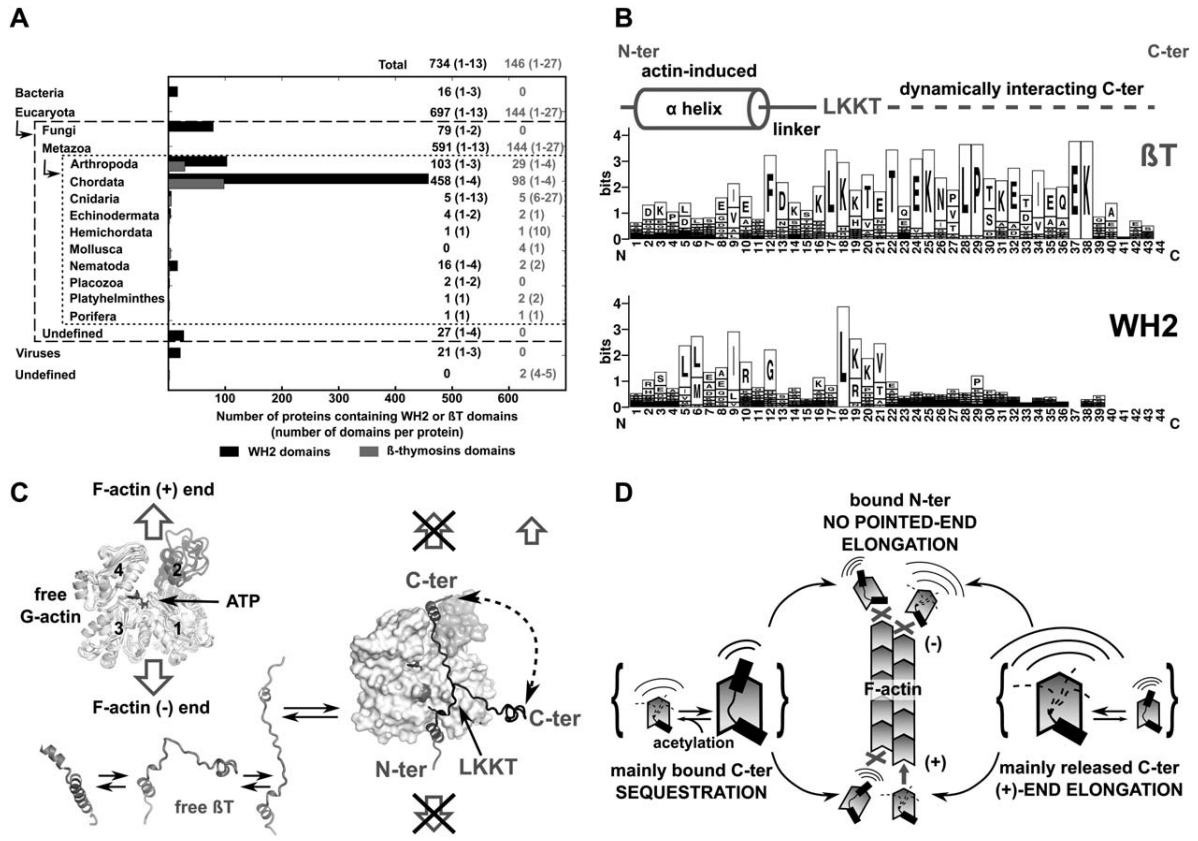


Fig. 1. Hallmarks and distribution of ßT and WH2 domains in proteins, and structural basis for their functional versatility as single domains bound to actin monomers. **(A)** Taxonomic distribution of proteins with ßT (gray labeling) or WH2 (black) domains in the SMART (Simple Modular Architecture Research Tool) database [Letunic et al., 2012]. The minimum and maximum number of repeats per protein found within each species is indicated in parenthesis. **(B)** The sequence homology of 30 ßT and 30 WH2 sequences found in proteins from different species is displayed as a sequence logo for ßT (upper) and WH2 (lower) sequence pattern [Crooks et al., 2004]. The height of each amino acid letter is proportional to their relative frequency at that position in the sequence. Above are indicated the main sequence elements and how they interact with G-actin. **(C)** Conformational variability of free G-actin, free and actin-bound βT. The latter correspond to full SAXS models of G-actin-bound CibD1 βT obtained either at low (sequestration function) or near physiological (profilin-like function) ionic strength [Didry et al., 2012]. Actin subdomains 1 and 3 (G-actin barbed face), and 2 and 4 (G-actin pointed face) are indicated on free G-actin. The intrinsic conformational flexibility of free G-actin is shown by four overlaid G-actin conformations (PDB: 1HLU, 2VCP, 2PDB or SAXS model [Didry et al., 2012]), and the flexibility of free βT by the solution structure of free Tß9, which adopts a mostly helical folding in 40% 1,1,1,3,3,3-hexafluoro-2-propanol (HFP) [Stoll et al., 1997], and different SAXS conformations of bound CibD1 βT, shown here without G-actin [Didry et al., 2012]. Arrows indicate to which end of the filament the monomer can associate with its barbed and pointed faces. **(D)** Structural basis for the functional versatility of single WH2/ßT domains in actin assembly. Functionally different βT/WH2 domains differ mainly by distinct dynamics of their C-terminal half interactions with G-actin pointed face that inhibit or allow association with the most dynamic barbed-end (+) of filaments and can be controlled by a single residue in long WH2/ßT [Didry et al., 2012]. In contrast, static interactions between WH2/ßT N-terminal half and G-actin barbed face inhibit always association with filament pointed-end (-).

tumours. Full-length or truncated Tß4 acts also extracellularly with less understood mechanisms, for example to promote dermal and corneal wound healing, or stimulate coronary vasculogenesis [Sosne et al., 2010]. WH2 domains were initially found in NPFs of the WASP and Scar/WAVE family that activate the Arp2/3 complex to generate branching networks of filaments in many actin-based motile processes involving membrane deformations (detailed below). WH2/ßT domains have been since identified or predicted as single or multiple units in a large number of modular signaling/regulatory proteins. The SMART database predicts that almost 850 proteins in eukaryotes and few dozen in prokaryotes or viruses contain a variable number of possible WH2/ßT domains (Fig. 1A). Like many IDRs, the latter display highly variable sequences and a weak sequence signature, which makes them difficult to identify or delimit properly (Fig. 1B). Their sequences vary between 25 and 55 amino acids [Paunola et al., 2002; Chereau et al., 2005; Carlier et al., 2011]. It is mainly identified by a central consensus actin-binding motif LKKT/V flanked by variable N-terminal and C-terminal extensions, with a N-terminal segment of about 8–12 residues predicted to fold into an

amphipatic α-helix of 2 to 3 turns (Fig. 1B). ßT share a more extended and conserved C-terminal half than WH2 (Fig. 1B). Sequence identity within the separated ßT and WH2 domain families is ∼40 and ∼20%, respectively, and drops to less than 15% when comparing sequences across the families. In the last years, WH2/ßT domains/repeats have emerged as multifunctional regulators of actin assembly dynamics but their exact contribution and regulatory molecular mechanism in most multidomain ABPs remain poorly understood or are still controversial [Dominguez, 2009; Campellone and Welch, 2010; Carlier et al., 2011].

The demonstration that individual WH2/ßT domains are archetypes of functional IDPs were first provided by NMR studies [Safer et al., 1997; Domanski et al., 2004]. NMR spectroscopy represents indeed the most powerful structural methods to analyze in solution the folding, conformational plasticity and dynamics of IDPs/IDRs on a large number of time scales in their native or bound states, or in a cellular crowded environment [McNulty et al., 2006; Bodart et al., 2008]. For IDRs involved in actin assembly regulation, detailed NMR analysis of actin conformations, and full characterization of complex interfaces are however limited by the quite large size of the system, the unavailability of (^{15}N, ^{13}C, ^{2}H)-labelled actin, which requires to be expressed in eukaryotic systems, and the difficulty to avoid polymerization of actin at the concentrations needed (typically 300 μM) for a full NMR study. We illustrate hereafter how the use of engineered chimeric proteins, actin-binding and polymerization assays, crystallographic, NMR, and SAXS (Small Angle X-ray Scattering) structural approaches and molecular dynamic simulations complement each other to decipher the molecular regulatory mechanisms and functional versatility of IDPs/IDRs in actin assembly dynamics.

Functional Versatility of Individual WH2/ßT IDRs in Actin Assembly Controlled by Fuzzy Complexes with G-actin

In solution, individual WH2/ßT domains characterized so far as isolated domains, bind usually preferentially G-actin-ATP over G-actin-ADP, inhibit the nucleotide exchange in G-actin, and do not or weakly interact with filaments [Carlier et al., 1993; Carlier et al., 1996; De La Cruz et al., 2000; Marchand et al., 2001; Mattila et al., 2003]. Their binding affinity for G-actin-ATP can be very sensitive to the ionic strength but near physiological ionic conditions they usually form complexes of moderate affinity with equilibrium dissociation constants (Kd) in the μM range [De La Cruz et al., 2000; Didry et al., 2012]. Their complexes with actin monomers inhibit the monomer association to filament pointed ends, but can either associate or not with the most dynamic barbed-ends of filaments to promote or inhibit their elongation. ß-thymosins (5 kDa) like Tß4 sequester G-actin, ie their complexes can not polymerize at either F-actin end. Most other WH2/ßT studied as individual domains in actin assembly, such as WASP and Scar1/WAVE1 WH2 domains [Egile et al., 1999; Higgs and Pollard, 1999], the 3 ßT repeats *of Drosophila* Ciboulot or its isolated first ßT domain (CibD1) [Hertzog et al., 2002], enhance motility by directing polarized assembly, with variable association rate constants for the binding of their complexes to barbed-ends. Their complexes behave for filament elongation as functional homolog of G-actin-ATP:profilin complexes. Their sequestering or profilin-like functions in actin assembly do not rely on different binding affinities for G-actin [De La Cruz et al., 2000; Hertzog et al., 2002; Didry et al., 2012] or obvious differences in their sequence composition. Many biochemical and structural studies have been required to understand at the molecular level the interactions of WH2/ßT with G-actin [Husson et al., 2010] and how the functional versatility between WH2/ßT is coded in their highly variable sequence and achieved at a mechanistic level in their complexes.

Circular dichroism data and NMR experiments using isotopically labeled IDRs have first shown that Tß4, Tß9 or Scar1 isolated WH2 are predominantly unfolded in their free native state in physiological buffers [Safer et al., 1997; Stoll et al., 1997; Domanski et al., 2004; Kelly et al., 2006]. Upon binding to G-actin-ATP, Tß4 folds completely. Its secondary structure elements that are otherwise transient are stabilized with G-actin and it displays a central extended region flanked by two N- and C-terminal helices [Safer et al., 1997; Simenel et al., 2000; Domanski et al., 2004]. To complement these structural studies, many crystal structures of various WH2/ßT domains bound to G-actin have been obtained using different polymerization inhibiting methods [Hertzog et al., 2004; Chereau et al., 2005; Aguda et al., 2006; Lee et al., 2007; Ducka et al., 2010; Rebowski et al., 2010; Didry et al., 2012]. These crystal structures show that highly variable WH2/ßT domains display a similar overall extended fold and binding path on G-actin with their N-terminal half but elusive interactions after their central LKKT/V motif. Their N-terminal amphipathic helix always caps the barbed face of G-actin by binding into the hydrophobic cleft between its subdomains 1 and 3, thus preventing association with the slow-growing pointed ends of filaments (Figs. 1C and 3E). Significant interactions with the pointed face of G-actin were observed in a crystal structure only while using a chimera of a C-terminal segment of Tß4 fused to gelsolin segment 1 in complex with G-actin [Irobi et al., 2004]. In this hybrid complex, Tß4 C-terminal residues (amino acid 30–39) fold into an α-helix that caps the pointed face of G-actin by binding between its subdomains 2 and 4. A full sequestering structural model of Tß4 was therefore proposed to block/cap both the barbed and pointed face of G-actin with Tß4 N- and C-terminal helices, respectively. WH2/ßT domains with profilin-like function were proposed in contrast to conserve G-actin pointed face free in complexes for association with filament barbed-ends. The

Table I. Multiple Interactions and Roles of IDRs in Proteins Representative of Different Activities in Actin Self-Assembly Dynamics

Proteins	IDR(s)	Functions related to the IDR(s)	Role of disorder	Refs
Tß4, *d. melanogaster* Ciboulot, WASP, WAVE, MIM, WIP, VASP,. . .	Single G-actin-binding WH2/ßT	sequester G-actin-ATP or promote its unidirectional assembly on barbed-end (profilin-like activity)	Fuzzy complexes: variable local dynamics in G-actin-bound states control variable association properties with barbed-end and can be tuned with a point mutation/modification	[Didry et al., 2012]
NPFs: WASP, N-WASP, WAVE, . . .	Middle B-GBD or meander-B regions, C-terminal V/WH2-CA	B-GBD: binding to $PIP_{2/3}$, Rho-GTPase, kinase/SH2, (V)CA VCA: binding to B-GBD, G-actin-ATP, Arp2/3	signal integrator, molecular hub, mutually exclusive interactions, cooperative interdomain middle-to-tail autoinhibition, cooperative allosteric activations	[Kim et al., 2000; Chen et al., 2010]
mDia1–3, INF2, FMNL3, or FMN1–2 subfamilies of formins	C-terminal DAD, WH2-DAD or FSI	N-term. DID binding, cooperativity/synergy for FH2-mediated G-actin nucleation or FH1-FH2-mediated filament elongation, WH2-DAD: filament severing in association with FH2, FSI: Spire-KIND binding	signal integrator, mutually exclusive interactions, interdomain head-to-tail autoinhibition, dynamic binding with G- and/or F-actin, FSI: dynamic tuning of FH2 activity with Spire	[Li and Higgs, 2005; Chhabra et al., 2009; Gould et al., 2011; Vizcarra et al., 2011; Bor et al., 2012]
CP (α:β heterodimer)	β-subunit C-term. tentacle (β-tentacle)	Folding & binding to barbed-end terminal actin required for tight capping of CP	optimal and dynamic binding of CP α:β heterodimer to filament barbed end	[Takeda et al., 2010]
CD2AP, CKIP, CARMIL,. . .	CARMIL-peptide	uncap CP α:β heterodimer bound to barbed-end	allosteric mechanism disturbing CP α:β bending/twisting bound to barbed-end	[Takeda et al., 2011; Kim et al., 2012]
Formins (FH1-FH2 domains)	FH1 domain: repeats of proline-rich motifs binding to profilin	Recruit and feed the FH2-bound barbed end with profilin:G-actin-ATP	FH1 disorder state offers a larger capture radius for profilin:G-actin, and with its non-redundant repeats, speeds the elongation rate of barbed-ends associated with FH2-Cterminal domains	[Shoemaker et al., 2000; Paul and Pollard, 2008; Courtemanche and Pollard, 2012]
Spire, Cobl, VopF/VopL, JMY, N-WASP	WH2/ßT repeats	sequester or nucleate G-actin-ATP/ADP, or promote G-actin unidirectional assembly, sever filament, cap barbed-end,. . .N-WASP, JMY: increase efficiency of branching mediated by NPF-Arp2/3	cooperative & dynamic binding with different actin tertiary & quaternary structures, repeats promote interactions with several actin subunits of filament or may assemble variable/dynamic G-actin oligomers	[Quinlan et al., 2005; Ahuja et al., 2007; Bosch et al., 2007; Husson et al., 2011; Gaucher et al., 2012]

TABLE I. *Continued*

Proteins	IDR(s)	Functions related to the IDR(s)	Role of disorder	Refs
JMY, MRTF-A or Phactr1 transcriptional cofactors	repeats of WH2/ßT or of RPEL-motifs	G-actin and importin-α binding, G-actin-concentration-sensing mechanism regulating nuclear localization for transcription	signal integrator, mutually exclusive interactions, cooperative G-actin binding	[Mouilleron et al., 2012; Zuchero et al., 2012]

WH2: WASP Homology 2 domain, βT: β-thymosin domain, NPFs: Nucleating Promoting Factors, FH1: Formin-Homology 1, B: polybasic region, GBD: GTPase-Binding Domain, VCA: Verprolin-homology/WH2, Central, Acidic regions, DAD: Dia autoregulatory domain, FSI: Formin-Spire Interacting, RPEL: actin-binding IDRs with core sequence RPxxxEL

switch of function of ßT domains from inhibition to promotion of actin assembly were expected to depend on different but illdefined interactions of their C-terminal region with actin subdomains 2 and 4, mediated by sequence variations widely distributed within their C-terminal half [Eadie et al., 2000; Hertzog et al., 2004; Irobi et al., 2004; Chereau et al., 2005; Aguda et al., 2006].

We have recently identified general structural basis for the sequestering and profilin-like function of ßT and WH2 in actin assembly [Didry et al., 2012]. The individual WH2/ßT of Tß4, CibD1, *Caenorhabditis elegans* Tetrathymosinβ containing 4 ßT domains [Van Troys et al., 2004], and human WASP-interacting protein (WIP) were used as models of functionally or structurally different WH2/ßT. Tß4 and CibD1 display more similar sequences but sequestering and profilin-like function, respectively, while Tß4 and WIP WH2 display only ∼15% sequence similarity. The segments in ßT that contribute the most to their overall affinity with G-actin and their function in actin assembly were determined using different chimeras of CibD1 ßT and Tβ4. We combined mutational analysis of the domains with actin binding and polymerization assays and several structural approaches, including protein crystallography, SAXS and NMR. Crystal structures were useful to analyze at high resolution all static/stable interactions in actin:ßT complexes, including those that affect distantly the behavior of disordered regions in complexes. SAXS and NMR have been instrumental to analyse the conformational dynamics of ßT IDRs in physiological buffers, and to correlate the structural changes in their complexes with functional differences measured in biochemical assays. The low resolution SAXS structural technique revealed overall conformational changes in complexes between low (non polymerizing G-buffer) and physiological (polymerizing F-buffer) ionic strength that correspond to different functions in actin assembly. The average conformations of flexible regions in complexes were modeled using the solution SAXS patterns of the complexes based on the uncomplete crystal structures of the complexes (Fig. 1C). NMR experiments using isotopically labeled ßT IDRs were essential to analyze in detail at different ionic strengths the folding and conformational dynamics of ßT domains between their free and actin bound states. Altogether the structural analyses demonstrate that the entire sequence of all ßT can interact with G-actin but with N-terminal static and C-terminal dynamic interactions. Single ßT domains with different functions do not need to target alternative actin binding sites via specific sequence variations in their central and C-terminal regions as previously thought, but to exhibit alternative dynamics of their C-terminal half interactions with G-actin pointed face (Figs. 1C–D). At physiological ionic strength these local interaction dynamics are primarily controlled by strong electrostatic interactions of a single residue along their sequence. In Tß4, this key residue is Lys14 that is located in the linker region between the N-terminal

amphipatic helix and the central consensus LKKT/V motif and provides a salt bridge with the highly conserved Glu334 of actin subdomain 3. Thus, a single salt bridge with G-actin located just upstream (in Tß4/10/15 ßT) or downstream (in WIP WH2) of their LKKT/V motif can induce a sequestering function in actin assembly with different and distantly related ßT or WH2 sequences. The removal of this strong electrostatic interaction between

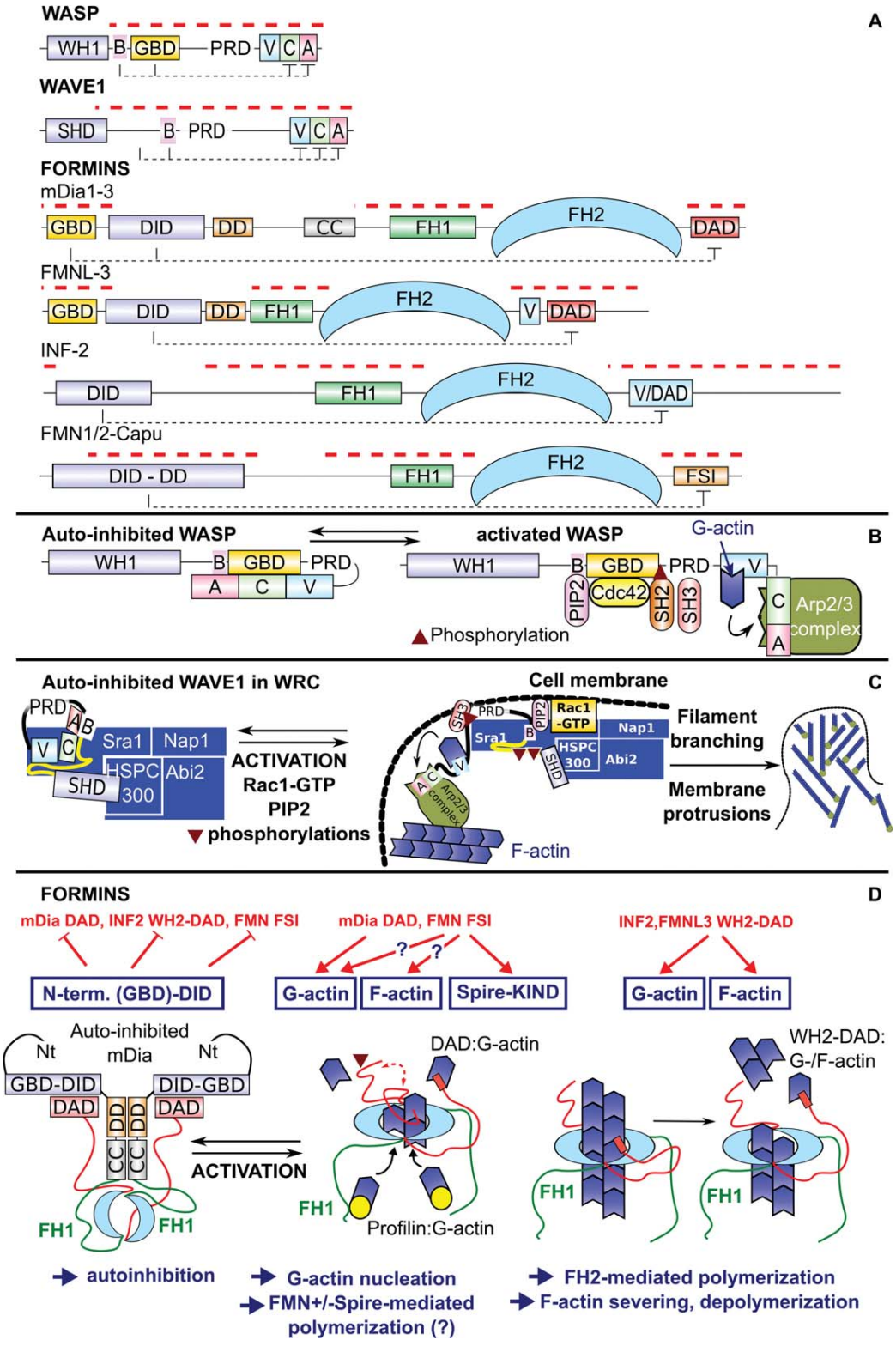


Fig. 2.

WH2/ßT and G-actin by a point mutation in Tß4 ßT or WIP WH2 on their key residue (K14Q/A in Tß4, R54N in WIP) is sufficient to reverse their function into a profilin-like function in actin assembly or to weaken their interaction with G-actin so much that it abolishes their primary function in actin assembly. Similarly, CibD1 ßT can be switched into an efficient sequestering ßT following a single point mutation that introduces a salt bridge with G-actin next to the LKKT/V motif like in Tß4 ßT. These results open perspectives for elucidating the multiple functions of both ßT and WH2 domains in other modular proteins [Didry et al., 2012].

IDR Regulation by Postranslational Modifications

WH2/ßT:G-actin complexes therefore represent fuzzy complexes in which the inherent disorder of WH2/ßT is partially conserved and functional in actin assembly (Table I). These results expand the variety of molecular mechanisms by which IDPs can fulfill different functions following small sequence modifications that affect their electrostatic interactions in complexes [Dyson and Wright, 2005; Tompa et al., 2005]. IDPs involved in cell signaling and transcription regulation recruit structurally dissimilar proteins and regulate different activities by adopting alternative conformations within their complexes. These alternative conformations are driven by the inherent structural plasticity of IDPs combined with the remodeling of their primary structure by reversible phosphorylations [Fabrega et al., 2003; Dyson and Wright, 2005; Tompa et al., 2005]. Here, the intrinsic structural disorder of WH2/ßT domains leads to a novel mode of functional versatility, in which the disordered protein does not regulate multiple targets, but induces different functions with the same target by varying the dynamics of their C-terminal interactions with small sequence differences. The activity of these IDRs in actin assembly may therefore be switched by single reversible sequence modifications like by other IDPs. In that regard ß-Thymosins undergo posttranslational modifications whose biological relevance is mostly unknown [Hannappel, 2010]. In a human proteomic study a fraction of Tß4 and Tß10 was found to be acetylated on the ϵ-amino group of Lys14 [Choudhary et al., 2009; Hannappel, 2010]. By abolishing the salt bridge with Glu334, acetylation of Lys14 can be predicted to decrease significantly the affinity for G-actin [Didry et al., 2012] and to switch the sequestering factors Tß4/10 into factors promoting unidirectional actin assembly at filament barbed-ends. Future experiments are required to challenge this possibility and to understand better how posttranslational modifications may impact *in vivo* the functions of WH2/ßT IDRs. Single phosphorylations play important roles in the regulation of many other unrelated IDRs found in multidomain ABPs. This is illustrated in the following section with recent molecular insights into the regulation of modular ABPs that accelerate *de novo* filament nucleation, a key activity in actin assembly.

Interfacial Properties of Individual IDRs Functionally Coupled with Other Neighbouring Domains in Modular ABPs

In multidomain proteins, IDRs are involved in many macromolecular interfaces and work in association with other adjacent disordered or folded domains. IDRs associated with the regulation of nucleating or CPs illustrate well how these segments control and coordinate multiple labile interactions at interdomain and protein–protein interfaces in modular ABPs.

Actin nucleation is the key rate-limiting step in spontaneous actin polymerization, and tight regulation of this process is critical to ensure that actin filaments form rapildy at the right time and place in cells. To date, three main different protein machineries have been proposed to nucleate new actin filaments: Arp2/3 complex activated by NPFs of the WASP/Scar/WAVE family, formin with profilin, and proteins containing nucleating WH2 repeats [Dominguez, 2009; Campellone and Welch, 2010]. The main relevant *in*

Fig. 2. Domain organization of WASP, WAVE1 and different formin subfamilies, and structural models describing how their structured or intrinsically disordered domains regulate the protein inactive and active states. (**A**) Modular domain architecture of human WASP, WAVE1 and different formins. Partially/fully IDR (dashed red lines), and autoinhibiting intramolecular interactions (dashed black lines) are indicated above and below each protein, respectively. A, acidic region; B, poly-basic region; C, central region; CC, coil-coiled region; DAD, Diaphanous autoinhibitory-like domain; DD, Dimerization-domain; DID, Diaphanous inhibitory-like domain; FH1/2, Formin-homology 1/2 domain; FSI, formin–Spire interaction domain; GBD, GTPase-binding domain; PIP2, PtdIns(4,5)P2, PtdIns-4,5-bisphosphate; PRD, Proline-rich domain; SH2/3, Src –homology 2/3 domain; SHD, Scar/WAVE homology domain; V, WH2; WH1, WASP homology domain 1; (**B**) In WASP autoinhibited state, its middle B-GBD and C-terminal C-A IDRs fold and bind with each other to make inaccessible the V-C-A regions to G-actin and Arp2/3. (**C**) WAVE1/2 is autoinhibited by intra- and intermolecular interactions within a complex of 5 proteins (Sra1, Nap1, HSPC300, Abi2 and WAVE1/2). In inactive WAVE1, its middle meander-B IDRs (yellow trace), located between its N-terminal structured SHD and unfolded PRD, integrate all the functional intramolecular inhibitory contacts with its C-terminal V-C-A IDRs. (**D**) In mDia, INF2, FMNL-3 or FMN subfamily of formin, the homodimeric G-/F-actin-binding FH2 domain is surrounded in each protomer by a FH1 and C-terminal region that are both at least partially disordered. Their C-terminal IDRs control both interdomain head-to-tail autoinhibition (DAD/FSI with N-term. GBD-DID/DID-like) and different interactions in the active state of formin in actin assembly (with G- and/or F-actin, or Spire KIND).

vivo functions of the latter remain however complex and debated as we will see in the last section [Renault et al., 2008; Qualmann and Kessels, 2009; Carlier et al., 2011]. All these machineries use different combinations of folded and disordered domains/proteins, and supposedly different molecular mechanisms of nucleation.

IDRs Found in NPFs of the WASP/Scar/WAVE Family Regulate Labile Cooperative and Antagonist Interactions

NPFs of the WASP/Scar/WAVE family are variable multidomain ABPs predicted to be extensively disordered (Fig. 2A) [Guharoy et al., 2013]. Working as membrane-bound stimulus responsive ABPs, they elicit cycles of filament branching that generate a dendritic array of actin filaments in many processes involving membrane deformations, such as lamellipodia protrusions in migrating cells, phagocytosis, dendritic spine activity, spatial organization of the Golgi or pathogen infections [Campellone and Welch, 2010; Rottner et al., 2010]. The prevailing model of regulation of WASP/Scar/WAVE ABPs is as follows. They exist in cells in conformational equilibria between inactive and active states. They are recruited on membrane and/or activated by phosphoinositide lipids (PIP_2 or PIP_3), phosphorylations, SH2 or SH3 (Src Homology 2/3) domain-containing proteins, or the active membrane-associated GTP-bound form of Rho small GTPases (Cdc42, Rac1,. . .). This leads to the disruption of their autoinhibited conformation in an individual or cooperative manner, and make accessible their conserved C-terminal constitutively active moiety, conventionally called V-C-A (Figs. 2B and 2C) [Campellone and Welch, 2010; Padrick and Rosen, 2010]. The C-terminal V-C-A region is composed of at least 3 IDRs of about 20–35 amino acids that are unfolded by themselves [Panchal et al., 2003; Kelly et al., 2006]: one to three WH2 domains (WH2/V), followed by a connector/central domain (C) and an acidic short extension (A). In this V-C-A region, each WH2/V domain can bind one G-actin molecule and is expected to deliver it to the actin-related Arp2 and/or Arp3 subunits of the 220kDa Arp2/3 complex, while the C-terminal CA domains bind one molecule of Arp2/3 complex (Fig. 2B). The resulting ternary V-C-A-actin-Arp2/3 complex associates with a preexisting filament after structural rearrangements in Arp2/3 complex to form the branched junction (Fig. 2C). Many of the molecular mechanisms involved in the branching reaction remain elusive, but few facets of the NPF IDR regulations have been highlighted at different stages of the process.

In WASP/N-WASP or WAVE proteins, the IDRs regulate different conformational equilibria between the NPF inactive and active states. The archetypal members of the family, WASP and N-WASP, are autoinhibited by interdomain interactions. Their full autoinhibitory interfaces remain unresolved but an important autoinhibitory middle-to-tail arrangement has been identified using a small domain fusion between the middle GTPase-binding domain (GBD) and C-terminal C region of WASP (Fig. 2B) [Kim et al., 2000]. The GBD IDR is like V-C-A largely unstructured by itself in its free state [Kim et al., 2000]. In this minimal reconstituted autoinhibited state of the NPF, WASP GBD and C region interact with each other to form a small compact folded assembly consisting of a short β-hairpin and five α-helices (Fig. 3A) that occludes partly Cdc42-GTP and Arp2/3 binding interfaces on each functional ID segments (Fig. 4) [Kim et al., 2000; Panchal et al., 2003; Hemsath et al., 2005; Kelly et al., 2006]. The resulting coupled folding and binding of the two distant GBD and C IDRs display a modest affinity corresponding to a Kd of ∼1μM [Peterson et al., 2004]. Activators relieve WASP/N-WASP autoinhibition allosterically by interacting with the NPF middle segments (B, GBD, and/or PRD IDRs, Fig. 2B), while the weak affinity of the C/CA:Arp2/3 interaction (Kd>200μM) is inefficient alone to compete with the autoinhibited GBD:C interface [Kelly et al., 2006]. Binding of Cdc42-GTP to the middle GBD has thus been observed to cause a dramatic conformational change of the inactive GBD structure, resulting in disruption of the hydrophobic core of the GBD:C contacts and release of the C-terminal C region (Figs. 3B, 3C, and 4) [Kim et al., 2000]. The GBD also acts as a signal integrator. Indeed, several WASP/N-WASP activators including PIP_2, Cdc42-GTP, and phosphorylation-mediated SH2 domain interact directly with or next to the GBD and activate WASP/N-WASP synergistically [Campellone and Welch, 2010; Padrick and Rosen, 2010]. A cluster of basic residues known as the polybasic (B) region is located at the GBD N-terminus (Fig. 2A) and provides an electrostatic steering mechanism of Cdc42-GTP recognition [Hemsath et al., 2005]. B is also essential for multiple PIP_2 headgroup binding and WASP/N-WASP membrane recruitment (Fig. 4). In the inactive state, the distant B and A regions are brought into close proximity by the GBD:C interaction. They may interact with each other in an ordered or disordered mode and extend the middle-to-tail interdomain autoinhibition to a larger B-GBD:C-A functional interface. NPFs are additionnally activated or inhibited following phosphorylations in response to various stimuli, associated for example with growth factors or substrate adhesion [Padrick and Rosen, 2010; Mendoza, 2013]. In WASP/N-WASP, these local modifications target residues located directly within their GBD or C IDRs or close to their autoinhibitory interface. Src-family kinases phosphorylate a conserved tyrosine sidechain within WASP/N-WASP GBD (Y291/Y256), and the casein kinase II (CKII) two serines at the end of their C region (S483S484/S484S485) (Fig. 4). WASP/N-WASP phosphorylations at Y291/Y256 activate the NPFs in a two-step process. In the autoinhibited GBD:C structure, Y291 is buried in one of the helix forming the five helical arrangement of the inactive interface formed by the two IDRs. After phosphorylation, a binding

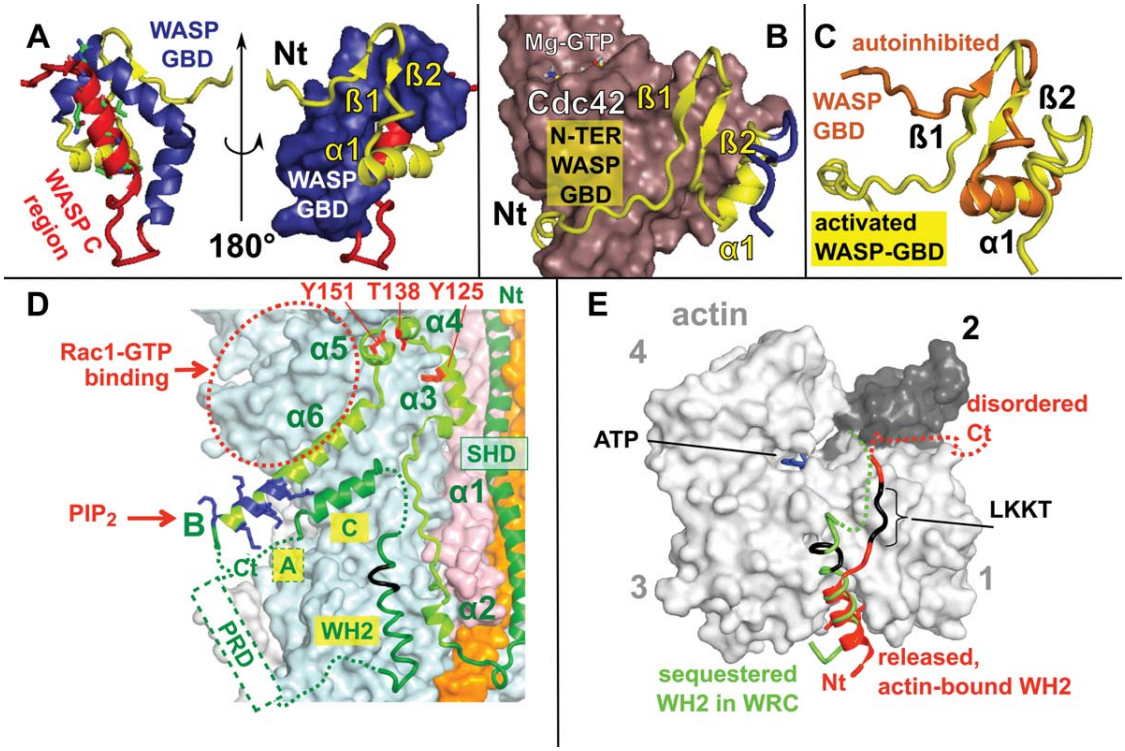


Fig. 3. Different known conformations of IDRs in WASP and WAVE1 inactive and active states and the mutually exclusive structural interfaces they control between the two states. (**A**) In WASP autoinhibited state, its middlle GBD (blue ribbon/surface and yellow ribbon) and C-terminal C (red ribbon) IDRs fold with each other to form a small compact folded assembly, which occludes partly Cdc42-GTP and Arp2/3 binding interfaces [Kim et al., 2000]. The side chains of C residues involved in hydrophobic and polar interactions at the interface are shown in green (L466, V467, L470, V473, Q475 and R477). (**B**) Activated, extended WASP-GBD (yellow/blue) bound to Cdc42-GTP (brown surface), with WASP-GBD shown in the same orientation as in the right panel of (A) [Abdul-Manan et al., 1999]. Only part of the C-terminal fragment of the GBD (in blue) is depicted since it unfolds upon binding of the GBD to Cdc42-GTP. (**C**) Superimposition of the two different conformations of WASP GBD bound to C (orange) or Cdc42-GTP (yellow). (**D**) The autoinhibited interface of WAVE1 (green ribbon) in the crystal structure of the minimal WAVE1 regulatory complex (WRC) [Chen et al., 2010]. The surfaces of Sra1, Nap1, HSPC300, and Abi2 are represented in cyan, white, pink, and orange, respectively. Disordered or missing (WAVE1 PRD) regions in the minimal WRC crystal structure are shown with green dashed lines. WAVE1 extended meander region (α2–α6) is shown in light green, and ordered basic residues of the B region in blue. The side chains of phosphorylation sites that are sufficient to induce allosterically the V-C-A release, leading to WRC activation, are shown in red. (**E**) The extended conformation of WAVE2 WH2 domain (in red) bound to G-actin-ATP (white surface) [Chereau et al., 2005]. Actin subdomains 1 to 4 are indicated. Disordered regions are shown with dashed lines. The alternative conformation of WAVE1 WH2 domain in its inhibited state within the WRC is superimposed in green with an orientation very close to the one displayed in (D). The two structural interfaces with either Sra1 (cyan surface in D) or G-actin (here) are mutually exclusive.

motif surrounding Y291 and including at least 5 to 7 residues [Huang et al., 2008] is expected to be recognized by the SH2 domains of Src-family kinases (Fig. 4). Binding of SH2 should promote and stabilize an extended conformation of the motif within the GBD, incompatible with the autoinhibited GBD:C interface. Altogether these data show that WASP/N-WASP IDRs can integrate multiple interactions (Figs. 2B and 4) and conformations (Figs. 3A–3C), and, consequently, control cooperative middle-to-tail autoinhibition and synergistic activation of the NPFs. These interfacial and regulatory properties of WASP/N-WASP IDRs seem to be conserved with Scar/WAVE/WASH IDRs.

Scar/WAVE/WASH proteins display a domain organization slightly different from WASP/N-WASP (Fig. 2A) and are autoinhibited within a pentameric WAVE/WASH regulatory complex (WRC) (Fig. 2C). The crystal structure of a minimal WRC that contains half of WAVE1 sequence has been recently solved in its inactive state, revealing a larger functional autoinhibitory interface of the NPFs than for WASP [Chen et al., 2010]. To solve by crystallography the pentameric and multidomain WRC of 380 kDa, Chen and coworkers reconstituted a functional minimal recombinant WRC optimized for structural homogeneity after truncations of WAVE1 PRD and Abi2 PRD-SH3 flexible regions. The region of WAVE1 between its N-terminal structured SHD (Scar/WAVE Homology Domain) and middle polybasic (B) region is predicted to contain significant disorder (Fig. 2A). It has been called the meander region because

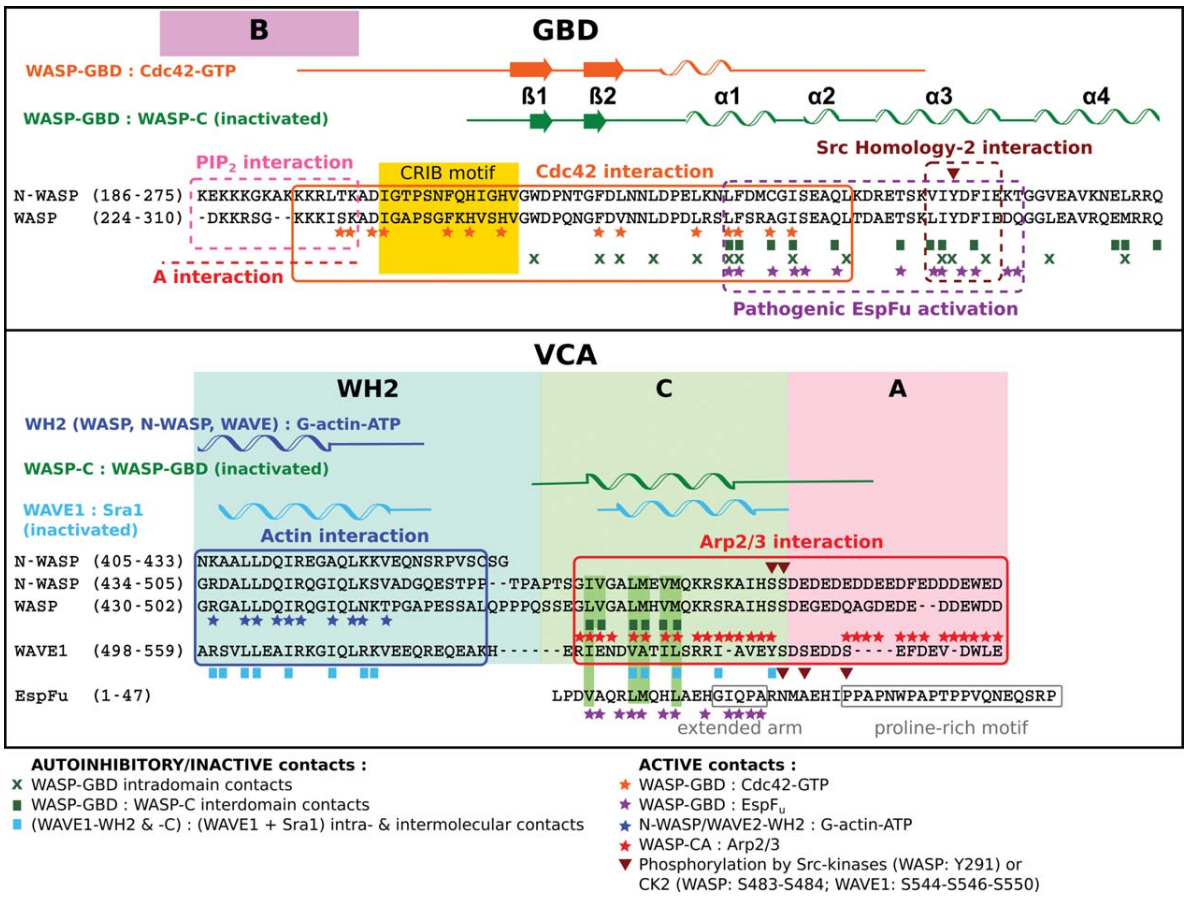


Fig. 4. The B-GBD IDRs of WASP act as a molecular node or hub in the multidomain NPF by integrating many interactions that are mutually exclusive or cooperative. The amino acid sequences of the B-GBD and V-C-A IDRs from human WASP, N-WASP and WAVE1, or of the N-WASP binding IDR of E. coli EspFu are shown with different labellings that underline residues deeply involved in involved in different interdomain and protein–protein interfaces. Secondary structures and interface residues are extracted from the solution or crystal structures obtained in [Abdul-Manan et al., 1999; Kim et al., 2000; Chereau et al., 2005; Cheng et al., 2008, 2010; Gaucher et al., 2012]. Only residues burying a solvent-accessible surface area > 30 Å^2 in the above structures are labeled. Arp2/3 interacting residues of WASP were identified by NMR line broadening experiments [Panchal et al., 2003]. The interaction boundaries that are not clearly delimited are shown in dashed lines. The B-GBD IDRs of WASP/N-WASP control 2 interdomain middle-to-tail autoinhibitory interactions, ie with C and possibly A IDRs (GBD:C and B:A), and 4 active interactions, ie with phosphoinositide lipids (B:PIP2), Cdc42-GTP (GBD:Cdc42-GTP), Src family kinases to phosphorylate Y291, and finally Src family SH2 domains after Y291 phosphosphorylation (GBD C-term.:SH2). The inactive and active interactions are mutually exclusive, while active B:PIP2, GBD:Cdc42-GTP and GBD C-term.:SH2 interactions are cooperative. In enterohaemorrhagic *E. coli* infection, the GBD IDR is targeted by the bacterial effector EspFu (GBD:EspFu) that outcompetes autoinhibited interactions and cellular activators upon a tight binding to GBD, leading to a kind of N-WASP constitutive activation.

within the inactive WRC, it adopts an extended, meandering folded conformation in contact of Sra1 protein. The meander region contains WAVE1 α-helices α2 to α6, with B starting at α6 C-terminus (Fig. 3D) [Chen et al., 2010]. The consecutive middle meander-B regions appear to coordinate many interactions in the WRC, like the B-GBD IDRs in WASP/N-WASP. In the inactive WRC structure, the C-terminal V-C-A region remains 40% disordered, with A being fully disordered. The V/WH2 domain of WAVE1 adopts an helical fold stabilized by both intra- and intermolecular interactions, with WAVE1 α2 helix and Sra1 protein, respectively (Fig. 3D). This interface occludes the WH2 residues which contribute most to G-actin interface (Figs. 3E and 4) [Chereau et al., 2005; Gaucher et al., 2012]. In parallel, the C region of WAVE1 folds into an amphipatic α-helix as in WASP GBD:C. This folding is also stabilized by both intramolecular interactions with WAVE1 α6 helix and intermolecular interactions with Sra1 protein (Fig. 3D). This interface masks C residues which have been shown by NMR to undergo a "disorder-to-order transition" upon Arp2/3 binding (Fig. 4) [Panchal et al., 2003; Kelly et al., 2006]. Finally, the disordered acidic (A) region at WAVE1 C-terminus is clearly positionned in the 3D-interface close to the middle phosphatidylinositol-binding B region (Figs. 2C and 3D). Ordered or disordered long range electrostatic contacts between the two distant

IDRs (B and A) may contribute to WAVE overall autoinhibition in the full-length WRC. Altogether the middle meander and B regions integrate therefore all the functional intramolecular inhibitory contacts with the C-terminal V-C-A (Fig. 2C). Interactions of the V-C-A, and perhaps the structural assembly of the entire meander:V-C-A element, appear also to be highly cooperative, as perturbation of either the V- or C-region contacts in Sra1 or in WAVE1 meander region produces WRC activity near that of the isolated V-C-A [Chen et al., 2010]. Mapping the binding area of WAVE activators on the inactive WRC interface shows additionally that the meander-B IDRs integrate or relay many allosteric signals that release the V-C-A contacts and activate synergistically the WRC. Rac1-GTP binds the inactive WRC with only a modest affinity (Kd ~1–2 μM) [Chen et al., 2010] and seems to cooperate with other factors for efficient membrane recruitment and activation of the WRC in actin assembly such as the small GTPase Arf1 [Koronakis et al., 2011] or phosphophoinositide lipids. Rac1-GTP binding area has been mapped on Sra1 in a patch directly adjacent to α4-α6 of WAVE1 meander region and possibly on WAVE1 α6 as well (Fig. 3D), and the binding of Rac1-GTP competes with the sequestration of V-C-A in WRC [Chen et al., 2010]. Interactions of Rac1 appear therefore to trigger conformational changes in the meander region and/or its contact site on Sra1. Similarly, the meander region is modulated by phosphorylations of Src, Abl kinases or Cdk5 at WAVE1 Tyr125 and Tyr151 and Thr138, respectively [Chen et al., 2010; Mendoza, 2013]. These residues are involved in inter- or intramolecular contacts of the meander inactive conformation in WRC but away from the V-C-A sequestering interface (Fig. 3D). A single phospho-mimetic mutant of either Tyr125 or Tyr151 in WRC is however sufficient to activate the complex both *in vitro* and *in vivo* [Chen et al., 2010]. Other phosphorylations on WAVE proteins that activate or inhibit the WRC [Mendoza, 2013] are localized on the neighbouring IDRs which are absent or disordered in the crystal structure, ie the PRD or A region. Finally, the conformation or stability of helix α6 C-terminus in the inactive WRC can be as well modified by the binding of phosphoinositide lipids to WAVE B region, the latter being essential for membrane recruitment of the WRC and formation of lamellipodia in cells [Oikawa et al., 2004]. Altogether the data suggest that the entire meander-B:V-C-A element functions as a cooperative labile folding block in WRC to control reversible interdomain middle-to-tail autoinhibition, and allosteric and synergistic activation, like the B-GBD:C-A IDR interface in WASP/N-WASP.

In conclusion, the IDRs of V-C-A-containing NPFs behave as efficient signal integrators. Their bindings of moderate affinity allow the control of multiple conformational equilibria that can be easily displaced by many partners in physiological or pathological processes. During bacterial intracellular infections, many bacteria subvert the actin cytoskeleton of host cells by hijacking Arp2/3 mediated nucleation. These pathogens inject either their own constitutively active V-C-A-containing proteins [Rottner et al., 2010], or unrelated proteins [Cheng et al., 2008; Aitio et al., 2012] or small molecules [Guenin-Mace et al., 2013] that disrupt the labile autoinhibited state of WASP/Scar/WAVE proteins. Enterohaemorrhagic *Escherichia coli* bacteria for example translocate EspFu into host cells during their intestinal colonization, to induce formation of actin-rich "pedestals" beneath the bacteria bound to human intestinal epithelial cells. EspFu contains 6 repeats of a 27-residue hydrophobic segment and a 20-residue proline-rich segment (Fig. 4) and is disordered in its native state [Aitio et al., 2012]. The N-terminal hydrophobic segment of each repeat activates N-WASP by mimicking structurally the autoinhibitory C:GBD interaction and competitively displacing C thanks to a higher affinity for GBD in the nanomolar range (Fig. 4) [Cheng et al., 2008]. The analysis of all the above structural interfaces involved in physiological or pathological processes show overall that the structural plasticity and dynamics of NPF IDRs allow to accommodate both autoinhibited compact and active extended conformations, and to drive variable and specific conformations with different targets. These alternative interactions can be efficiently regulated by single phosphorylations. The IDRs act as interaction hubs in these modular ABPs (Table I). In WASP/N-WASP for example, the continuous B and GBD IDRs coordinate six physiological and one pathological interfaces that are mutually exclusive or cooperative and regulate the NPF inactive and active states (Fig. 4). Additional interactions likely occurs, that will be uncovered when the NPF full autoinhibitory interfaces or more complete active interfaces will be analyzed. The short adjacent binding sites of physiological activators with B-GBD may work cooperatively by stabilizing altogether the 90 amino acid segment in an extended conformation, each additional binding making the conformational transition towards the B-GBD:C-A autoinhibitory compact folding less reversible. Future studies will delineate the conformational dynamics of these mutliple interfaces involving several IDRs, and how, after V-C-A release, the inherent flexibility of IDRs in the resulting ternary V-C-A-actin-Arp2/3 complex may be important as well for promoting all the dynamic molecular arrangements required in the branching process.

Multi-Functionality of IDRs Found in Formins at the C-terminus of Their Main Catalytic Actin-Binding FH2 Domain

Unlike Arp2/3 and VCA containing NPFs, formin proteins nucleate new filaments that are unbranched. They thus assemble diverse cellular structures, including the cytokinetic contractile ring, polarized actin cables, stress fibers and filopodia [Goode and Eck, 2007; Campellone and Welch, 2010]. Based on their different modular

architectures, formins have been classified into at least seven main subclasses in metazoans. The modular architecture of four representative subfamilies is reported in Fig. 2A. The basic catalytic units that identify formins in actin assembly is related with the presence of a formin homology 1 (FH1) and 2 (FH2) domain. Both are located in the C-terminal part of formins while other domains are more variable and adapted to different cell cues (Fig. 2A). Several formin subfamilies contain for example N-terminally a GBD IDR that contributes to their activation on membrane upon binding to activated Rho GTPases. FH2 domains form flexible donut-like dimers that can accelerate *de novo* filament nucleation from G-actins by stabilizing two G-actin between their arms in an arrangement mimicking the lateral actin:actin contacts of filaments and by providing smaller binding interfaces with two other monomers per FH2 protomer. FH2 dimers also bind to filament barbed-end and protect it from barbed-end CPs that stop elongation. In full-length proteins, FH2 dimers can translocate with the barbed-end as it elongates, acting as processive elongating factors. Additional high-resolution structural studies are required to understand the translocation of formins or the competition between formin and barbed-end CPs at the most dynamic end of actin filaments. The crystal structures of G-actin-bound FH2 dimers [Otomo et al., 2005; Thompson et al., 2013] and the cryo-electron microscopy model of the heterodimeric CP bound to F-actin barbed end [Narita et al., 2006] suggest that free filament barbed-ends cannot accommodate a simultaneous binding of formin and CP because filament-bound FH2 dimer or CP α/β heterodimer are expected to display overlapping binding sites for the terminal actin barbed-faces. Intrinsically disordered FH1 domains are composed of multiple profilin-binding polyproline segments. They catch and deliver G-actin:profilin complexes to the FH2-bound barbed-end to promote formin-mediated processive elongation of growing barbed-ends. Recent biochemical and structural data suggest that the catalytic machinery of different formin subclasses is tripartite because it requires as well IDRs located at the C-terminus of their FH2. These IDRs, which can be unrelated between formin subclasses, share several functional properties with the IDRs of V-C-A-containing NPFs by regulating multiple interactions and conformational transitions between formin inactive and active states (Table I).

IDRs at FH2 C-terminus regulate first interdomain head-to-tail autoinhibition. The two subclasses of Diaphanous-related formins (mDia1–3) and formin homology domain proteins (FHOD1/3) are activated by Rho family GTPases and contain a C-terminal Diaphanous autoregulatory domain (DAD) [Campellone and Welch, 2010] identified as an IDR in disorder prediction server (Fig. 2A). Their N-terminus contains a GBD and Diaphanous inhibitory domain (DID) that participate in autoinhibition. Their C-terminal DAD IDR binds the GBD–DID to inhibit the actin polymerizing activity of the FH2 domain (Fig. 2D, left panel) [Wallar et al., 2006]. This inhibitory interaction is disrupted by the binding of Rho GTPases to the GBD–DID [Lammers et al., 2005] and results in activation of the C-teminal FH2 in vitro [Li and Higgs, 2005]. Structural comparisons of mDia GBD–DID bound to either a DAD or to Rho GTPases indicate that binding of Rho and DAD is mutually exclusive, and their binding sites in the GBD–DID partially overlap explaining the molecular basis for the allosteric activation [Campellone and Welch, 2010]. Importantly, mDia activation on GTPase binding in vitro is incomplete, raising the possibility that additional cellular factors may coalesce upon the autoinhibited GBD-DID interface for full activation like in V-C-A-containing NPFs. In that matter, the DAD of FHOD1 or mDia2 is phosphorylated by ROCK (rho-kinase). This relieves the formin autoinhibition *in vitro* and promotes FHOD1-mediated stress fibre assembly in cells [Takeya et al., 2008], or mDia2-mediated actin polymerization and smooth muscle cell-specific gene transcription [Staus et al., 2011]. Founding mammalian formins (FMN1/2 or *Drosophila melanogaster* Cappucino (Capu)), FMNL-3 or INF2 (Inverted Formins) formins display a different overall modular architecture, lacking an N-terminal GBD (except for FMNL-3) and ending either with an FSI (Formin-Spire Interacting) IDR for FMN, or with WH2-DAD IDRs for FMNL-3 and INF2 (Fig. 2A). Both FMN FSI [Bor et al., 2012] and INF2 WH2-DAD [Chhabra et al., 2009] IDRs at their FH1-FH2 C-terminus are involved in interdomain head-to-tail autoinhibition with an unrelated N-terminal DID.

In addition to their role in interdomain autoinhibition, these IDRs control as well multiple interactions in formin active states. They appear to be required for the catalytic nucleating and/or processive activity of their FH2 on filament barbed-ends, and are sometimes involved in alternative interactions with other ABPs (Fig. 2D). The DAD of mDia1 interacts with G-actin and works in concert with the FH2 to enhance nucleation without affecting the rate of filament elongation, and independently from the FH1 domain (Fig. 2D, middle panel) [Gould et al., 2011]. The WH2 and DAD containing C-terminus of INF2 [Chhabra and Higgs, 2006] or FMNL-3 [Heimsath and Higgs, 2012] act in synergy with their FH2 domain to accelerate actin polymerization from monomers. Additionally, the C-terminal region of INF2 is required for actin filament severing, while the C-terminal region of FMNL-3 enhances the otherwise modest severing activity of FMNL-3-(FH1-FH2) (Fig. 2D, right panel). The isolated C-terminus of FMNL-3 including its WH2-DAD IDRs binds both monomers and filament barbed-ends [Heimsath and Higgs, 2012]. Finally, the C-terminal FSI of FMN-subfamily formins cooperates with their FH2 domain to accelerate actin polymerization [Vizcarra et al., 2011], and interacts as well with the N-terminal KIND (Kinase homology domain) of Spire

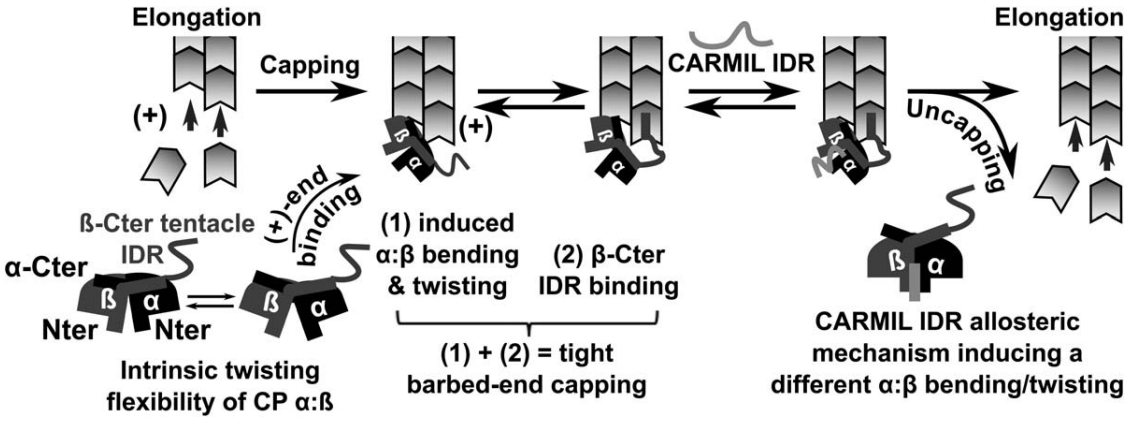


Fig. 5. Structural basis for the regulation of two unrelated IDRs associated with the capping activity of CP at filament barbed-end. The CP is an intrinsically twisting, pseudosymmetric α:β heterodimer with the shape of a mushroom, including a stalk containing α and β N-terminii and a cap containing their C-terminii. CP binds to F-actin barbed-end with its "mushroom cap" and the Cterminal IDR of the β-subunit (β-tentacle) is important in the dynamic and optimal binding to the filament barbed-end (step 2). In contrast, disordered CARMIL peptides perturb allosterically the overall optimal bending and twisting of CP heterodimer induced by its barbed-end binding. CARMIL IDRs bind to CP "mushroom stalk" across its two CP-L and CP-S twisting domains [Takeda et al., 2010; Kim et al., 2012].

[Vizcarra et al., 2011; Zeth et al., 2011]. The molecular bases for these multiple regulations in actin assembly are not yet understood. It remains to be determined if and how the C-terminal FSI interacts with G- and/or F-actin in presence of the FH2. The regulation with Spire is even more complex since Spire N-terminal half contains four consecutive WH2 domains next to its KIND domains that altogether display multiple activities in actin assembly with both G- and F-actin as described below. As isolated domain, the interaction of the N-terminal KIND of Spire with the FMN FSI inhibits the acceleration of actin polymerization induced by FMN FH2 and C-terminal domains, while a longer N-terminal construct of Spire that includes its KIND and 4 consecutive WH2 domains, enhances in contrast synergistically the FMN-induced actin polymerization [Quinlan et al., 2007; Vizcarra et al., 2011]. Future studies are required to elucidate the complex multidomain regulation and synergy between Spire and FMN nucleating factors at filament barbed end, which involve multiple IDRs in the two ABPs.

Adaptability and Allostery of IDRs Associated with the Capping Regulation of CP

Structurally unrelated CPs like the CP [Zwolak et al., 2010] or Eps8 [Hertzog et al., 2010] use C-terminal IDRs to regulate their barbed-end capping activity. The regulatory molecular mechanisms of CP have been lately more extensively characterized, including inhibitory mechanisms by disordered CARMIL peptides. CP facilitates the actin turnover in the formation of various actin structures, including Arp2/3-mediated dendritic arrays, by reducing the overall number of elongating barbed-ends. It is an α:β heterodimer. Its two folded subunits of ~30 kDa form a stable pseudosymmetric structure with the shape of a mushroom, with a stalk and cap (Fig. 5). NMR and crystallographic data have shown that the C-terminal ~30 amino acids of the *β*-subunit, called the *β*-tentacle, are disordered [Zwolak et al., 2010; Takeda et al., 2011]. CP binds to the barbed-end of actin filaments with high affinity (Kd ~0.1 nM) and terminates assembly and disassembly at the end, i.e. caps the filament.

Several structures of CP, including a low resolution EM model of CP bound to filament [Narita et al., 2006], are available to approach the different steps of the capping reaction and the role of local disorder in CP at the atomic scale (Table I). Structural comparisons and molecular dynamic simulations revealed that CP is composed of two rigid domains (CP-L and CP-S) that undergo an intrinsic twisting motion relative to each other but do not directly correspond to the subunit interface [Takeda et al., 2010; Kim et al., 2012]. The current models describe the barbed-end capping as following [Narita et al., 2006; Kim et al., 2010; Takeda et al., 2010, 2011; Kim et al., 2012]. CP first associates with filament barbed-end by the cap surface involving significantly the α-subunit C-terminus (Fig. 5). CP then firmly binds to the barbed-end by two independent processes. (1) The interaction with filament barbed-end induces a particular bending and twisting motion of the cap. (2) The β-tentacle IDR can eventually reach the hydrophobic cleft between subdomain 1 and 3 of the terminal actin in order to fold and bind to the barbed-end. In this interface, the IDR is therefore important for a dynamic and optimal binding to the barbed-end. Its deletion induces a two-fold reduction of CP affinity for barbed-ends [Kim et al., 2010]. Local disorder in the β subunit C-terminus may as well be important to accommodate the conceivable conformational dynamics of the two last actin subunits at barbed-ends. In that regard all folded proteins that cap filament barbed-ends such as Gelsolin, twinfilin, or formin, are similarly

composed of folded domains including or connected by flexible or disordered linker/regions.

Recent structural data on CP bound to several inhibitory CARMIL peptides illustrate further how disorder participates as well in the allosteric inhibition of CP. The modular CARMIL protein facilitates the dissociation of CP from the barbed-end (uncapping activity) mainly via a conserved short CP-binding motif (CPB) of ~20 amino acids present in the otherwise unrelated proteins like CD2AP and CKIP-1. The CPB is surrounded by flanking regions that contribute to the overall binding affinity to CP and extend the interactions more significantly on the peptide C-terminal side. These CARMIL peptides are intrinsically disordered. Several crystal structures show that CARMIL peptides wrap around the stalk of the mushroom-shaped CP at a site distant from the actin-binding interface on the top of the mushroom cap (Fig. 5) [Hernandez-Valladares et al., 2010; Takeda et al., 2010, 2011]. Molecular dynamic simulations suggest that CARMIL peptides inhibit CP allosterically by altering the conformational twisting flexibility between CP-L and CP-S from the mushroom stalk binding [Takeda et al., 2010, 2011; Kim et al., 2012]. The CPB-binding is sufficient to induce this allosteric mechanism, while the C-terminal flanking region in CD2AP and CARMIL appear only to increase the overall peptide stability and affinity with CP, suggesting a decoupling between the regions responsible for binding specificity (stronger contribution of flanking regions than CPB) and function (e.g., CPB) [Takeda et al., 2011]. The flexible conformation of the CARMIL peptides was also proposed to be potentially important for intrusion into the mobile binding groove across the two twisting rigid domains CP-L and CP-S [Takeda et al., 2011].

Complex Multifunctional Regulations of Intrinsically Disordered Repeats in ABPs

The preponderance of repeats in intrinsically disordered sequences is high and may represent a way to expand IDP functions in evolution [Tompa, 2003]. Repeats can be either functionally equivalent or nonredundant, or introduce novel functions that emerge as the repeat number exceeds a threshold level. Similarly, many ABPs contain diverse modular organizations of repeated IDRs as emphasized for WH2/βT domains in Fig. 1A. Their molecular mechanisms of regulations remain largely poorly understood.

Nonredundant Repeats of FH1 Polyproline Stretches Work in Synergy in Formin to Speed the Elongation Rate of FH2-Associated Barbed-Ends

In formin, the FH1 domain in presence of profilin speeds the elongation of individual barbed-ends associated with FH1-FH2-Cterminus domains, often above the diffusion-limited rate of elongation of free barbed-ends [Paul and Pollard, 2009]. FH1 domains of most formins lie N-terminal to their FH2 domain and are predicted to be mostly disordered except for multiple discrete stretches (often between 2 and 8) of contiguous proline residues, that can transiently form type-II polyproline helices. Each polyproline track displays individually weak affinity for free or actin-bound profilin, with Kd values varying from ~5 to 20 μM for stretches that include 13 to 20 prolines to >1000 μM for stretches including only 5 to 8 prolines [Perelroizen et al., 1994; Petrella et al., 1996; Courtemanche and Pollard, 2012]. Transfer of actin monomers to the barbed-end of actin filament is established by diffusion and closure of FH1 loops filled with profilin:G-actin on the FH2-bound barbed-end (Fig. 2D, middle panel). This enables direct addition of actin monomers to the FH2-bound barbed-end and may promote translocation of one of the FH2 protomer while the full FH2 dimer remains processively attached to the end. Recent data show that in presence of profilin, the rates of elongation of filaments associated with the FH1-FH2-Cterminus of yeast formin Bni1p increased with the total number of FH1 polyproline tracks [Paul and Pollard, 2008]. However, each FH1 polyproline track contributes differently and is nonredundant along the FH1 sequence [Paul and Pollard, 2008; Courtemanche and Pollard, 2012]. Their contribution to the barbed-end elongation rate vary according to their distance in the FH1 sequence to the FH2, and to their binding strength for profilin:actin. Their distance to the FH2 domain determines the volume explored by diffusing profilin-actin bound to a polyproline track before contacting the barbed-end. The affinity of each polyproline track in the FH1 domain of Bni1p and possibly of other formin for profilin-actin is tuned relative to its distance to the FH2 domain [Courtemanche and Pollard, 2012]. Near the FH2 domain, FH1 polyproline tracks contain a smaller number of prolines in order to display lower affinity for profilin, which favors rapid and dynamic G-actin transfer to the barbed-end upon FH1 loop closure. Far from the FH2, they contain in contrast a larger number of prolines to display higher affinity for profilin (slower dissociation rates) and to compensate for the slower loop closure rates. In conclusion, the individual weak affinity of these IDRs allows driving dynamic interactions at the elongating barbed-end. The FH1 disorder state may provide a larger capture radius for profilin:G-actin leading to faster on-rates for binding as it has been proposed for IDRs involved in DNA binding [Pontius, 1993; Shoemaker et al., 2000]. The repeats speed the overall elongation mediated by formin, and the specificity of each repeats along the FH1 sequence optimizes G-actin transfer efficiency with the number of repeats or the length of the flexible loop relative to its distance to the FH2 (Table I).

Non-Redundant Repeats of WH2/ßT IDRs Cooperate in Several ABPs to Regulate the Dynamics of G- and/or F-Actin Assembly/ Disassembly

Some modular organizations containing repeats of actin-binding WH2/ßT IDRs extend significantly the panoply of functions regulated by single WH2/ßT domains in actin assembly (Table I) [Renault et al., 2008; Husson et al., 2010; Carlier et al., 2011]. The specificities and synergy of multiple ßT domains were highlighted in C. elegans TetraThymosinß that contains 4 consecutive ßT domains and binds multiple G-actins *in vitro* [Van Troys et al., 2004]. Mutants with only one repeat intact or that have only one repeat rendered unfunctional or less active were analyzed in G/F-actin-binding and polymerization assays. Some repeats were thus shown to exhibit specificity for either G-actin (repeats 1 and 4), F-actin (repeat 3) binding or both (repeat 2). Together the four repeats cooperate to display the full biochemical activity of TetraThymosinß in G- and F-actin assembly regulation, including its desequestering ability in presence of Tß4 [Van Troys et al., 2004]. Spire [Quinlan et al., 2005], Cordon-bleu [Ahuja et al., 2007], V-C-A-containing JMY [Zuchero et al., 2009] or the pathogen VopF/VopL effectors [Liverman et al., 2007; Tam et al., 2007] contain different modular organizations with 3 to 4 WH2/ßT repeats that accelerate *de novo* filament nucleation from pure G-actin. They have been then proposed to represent a new subfamily of actin nucleators. The specificities that confer nucleation for these repeats but not in others are not yet well understood. In Spire, the 4 contiguous WH2 bind cooperatively 4 G-actin with a Kd of ~0.15 μM^4 [Bosch et al., 2007] but are functionally not redundant for nucleation. Spire last two WH2 are more determinant for nucleation than its first two domains [Quinlan et al., 2005]. In Cobl containing 3 WH2, nucleation of pure G-actin relies mainly on the combination of its first WH2 associated with a predicted disordered Lys-rich region, located directly N-terminal to the WH2 [Husson et al., 2011]. Several observations suggest that nucleation may not represent the primary *in vivo* function of these repeats [Renault et al., 2008; Carlier et al., 2011]. First, compared to other known actin nucleators like formin, WH2 repeats of Spire or Cobl display a relatively weak nucleating efficiency *in vitro* in view of their affinity for G-actin, suggesting that only a very small portion of their complexes with G-actin monomers behave as nuclei. Their complexes appear to exist in solution in equilibrium with other main functional states that regulate other activities such as profilin-like and/or sequestering activities, including sequestration of G-actin-ADP after filament depolymerization [Bosch et al., 2007; Husson et al., 2011]. Secondly, the nucleating activity of Spire or Cobl WH2 repeats is abolished *in vitro* when G-actin is bound to profilin [Bosch et al., 2007; Husson et al., 2011] and *in vivo*, Spire regulates actin dynamics in many developmental processes in concert with FMN-subfamily formin and profilin [Manseau and Schupbach, 1989; Dahlgaard et al., 2007; Pfender et al., 2011]. An important clue to understand the role of WH2/ßT repeats in ABPs may therefore rely on their capacity to interact jointly with the side and/or barbed-end of filaments, which is otherwise weak with individual WH2. The 4 WH2 of Spire bind to filament barbed-ends with a good affinity, i.e. with an equilibrium dissociation constant of approximately 0.010 μM [Bosch et al., 2007]. The 4 WH2 repeats of Spire, or first two WH2 domains of Cobl associated with their N-terminal Lysine-rich motif, modulate both the dynamics of assembly and disassembly of filaments. They regulate filament severing with different efficiencies, and Spire regulates also barbed-end capping in presence of profilin [Bosch et al., 2007; Carlier et al., 2011; Husson et al., 2011]. Similarly, the presence of two consecutive WH2 in the V regions of N-WASP introduce a capacity to interact with filament barbed-end [Gaucher et al., 2012], and a higher branching efficiency in N-WASP V-C-A than in WASP or WAVE V-C-A [Yamaguchi et al., 2000]. Finally, in the V-C-A of JMY, the function of its three consecutive WH2 repeats is further extended by integrating an overlaping importin-α-binding nuclear-localizing sequence [Zuchero et al., 2012]. The mutually exclusive interactions between G-actin and importin-α/β bindings to the repeats regulate JMY cytoplasmic or nuclear localization, and consequently the nuclear transcription induced by the protein with G-actin cytoplasmic concentration and binding (Table I). This is similar to the regulation of RPEL-motifs that are other small, intrinsically disordered G-actin-binding domains present as repeats in several transcriptional co-factors such as MAL/MRTF-A or Phactr1 [Mouilleron et al., 2008, 2012]. In conclusion, the intrinsic conformational dynamics of these small actin binding domains allow them in repetition to interact cooperatively with multiple G-actin-ATP/ADP and/or with the side or barbed-end of filaments. They can thus act as versatile or multifunctional regulators of the G-actin pool, and/or of the dynamics of assembly and disassembly of filaments, regulating multiple key activities in actin assembly such as G-actin nucleation, unidirectional assembly or sequestration, filament severing or capping. What are the main activities relevant in vivo for each modular organizations with WH2/ßT repeats in presence of other ABPs, such as FMN-subfamily formin and profilin for Spire, remain major issues to be solved. These investigations will require to understand better what are the specificities at the level of each individual WH2/ßT, the synergy of their different associations, and the exact number of WH2/ßT which support each of their activities in actin assembly. For deciphering the molecular bases of their versatility or multifunctionality, it will be essential to analyze different conformations and the dynamics of their complexes which may include local disorder and fuzzy complexes like it has been observed at the level of their elementary functional unit with a single ßT bound to one G-actin [Didry et al., 2012].

Concluding Remarks

The structural and functional properties of IDRs involved in actin cytoskeletal remodeling begin only to emerge in the light of few complementary studies that dissect and approach their conformational dynamics in simplified molecular and functional models of multimodular ABPs and actin assembly dynamics. These studies suggest that actin cytoskeletal IDRs are especially designed to relay many labile interactions in ABPs (Table I).

IDRs of ABPs share important properties of IDRs involved in signal transduction or transcription regulation. Thanks to their inherent structural plasticity and dynamics, they can accommodate both autoinhibited compact and active extended conformations, and adopt variable conformations with different and structurally unrelated targets (Fig. 3). By adopting extended conformations, they can build highly specific interfaces with shorter sequences than structured globular domains [Gunasekaran et al., 2003]. The folding upon binding processes provides high specificity with relatively low affinity due to the entropic cost associated with the disorder-to-order transition [Dyson and Wright, 2005]. Small but highly specific functional segments are adapted for relaying reversible and transient interactions at macromolecular interfaces in actin self-assembly dynamics. Disorder may be partially preserved and functional in their bound states [Didry et al., 2012]. Intrinsic conformational flexibility facilitates access to enzymes that introduce or remove post-translational modifications. In bound states, intrinsic conformational dynamics and moderate binding strengths enable complete remodelings of macromolecular interfaces after single post-translational modifications like phosphorylations (WAVE1 meander region, WASP/N-WASP/WAVE1 C-A, formin FHoD1/mDia2 DAD,. . .) or possibly lysine acetylations (Tβ4/10).

The conformational variability and adaptability of IDRs and their regulations via diverse post-translational modifications allow them to act as interaction hubs in ABPs. Their intrinsically disordered sequence can integrate multiple overlapping binding sites of different domains and/or proteins and regulate many mutually exclusive interactions (Fig. 4). They thus regulate reversible switches between the inactive and active states of modular ABPs, as it has been highlighted here for V-C-A containing NPFs or different formin subfamilies (Fig. 2). In these ABPs, one or several distant IDRs control interdomain head/middle-to-tail autoinhibition and consequently allosteric activation. The antagonistic interactions induced by overlapping binding sites on IDRs may also be between different partners in the active states of ABPs (Figs. 2, 4, and Table I). In macromolecular interfaces, multiple consecutive IDRs or long IDRs that can integrate distinct nonoverlapping binding regions, regulate cooperative interactions with their targets. WASP/N-WASP B-GBD IDRs or WAVE1 meander-B IDRs integrate many activating signals/interactions that each modifies the conformational equilibria between the inactive and active states of the IDRs and thus synergize or cooperate to activate the NPF. Similarly, multiple WH2/βT domains in Spire or Cordon-Bleu introduce an efficient regulation of the dynamics of assembly and/or disassembly of filaments by a cooperative binding of several flexible actin-binding IDRs with several actin subunits of filaments [Bosch et al., 2007; Renault et al., 2008; Carlier et al., 2011; Husson et al., 2011]. Finally, IDRs display often poor sequence signature in correlation with the fact that their sequence contain less folding constraints than intrinsically structured domains. The multiple regulations we have highlighted here for IDRs in modular ABPs further explain why known functional IDRs like actin-binding WH2/βT display highly variable sequences between ABPs. By controlling the autoinhibited states of modular proteins and recognizing several partners in their active states, the functional sequence of each IDR is specifically engineered in each ABP to permit a fine-tuning of multiple intra- and intermolecular interactions. Overall, the numerous and labile interactions of IDRs in modular ABPs contribute to explain how multidomain proteins can integrate and coordinate efficiently multiple signals and interactions with high turnover in actin self-assembly dynamics. A great deal of work remains to be done to further elucidate the full panoply of IDR regulations, in a larger number of ABPs, and in more complete and integrated systems.

References

Abdul-Manan N, Aghazadeh B, Liu GA, Majumdar A, Ouerfelli O, Siminovitch KA, Rosen MK. 1999. Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. Nature 399(6734):379–383.

Aguda AH, Xue B, Irobi E, Preat T, Robinson RC. 2006. The structural basis of actin interaction with multiple WH2/beta-thymosin motif-containing proteins. Structure 14(3):469–476.

Ahuja R, Pinyol R, Reichenbach N, Custer L, Klingensmith J, Kessels MM, Qualmann B. 2007. Cordon-bleu is an actin nucleation factor and controls neuronal morphology. Cell 131(2):337–350.

Aitio O, Hellman M, Skehan B, Kesti T, Leong JM, Saksela K, Permi P. 2012. Enterohaemorrhagic Escherichia coli exploits a tryptophan switch to hijack host f-actin assembly. Structure 20(10): 1692–1703.

Bisi S, Disanza A, Malinverno C, Frittoli E, Palamidessi A, Scita G. 2013. Membrane and actin dynamics interplay at lamellipodia leading edge. Curr Opin Cell Biol 25(5):565–573.

Bodart JF, Wieruszeski JM, Amniai L, Leroy A, Landrieu I, Rousseau-Lescuyer A, Vilain JP, Lippens G. 2008. NMR observation of Tau in Xenopus oocytes. J Magn Reson 192(2):252–257.

Bor B, Vizcarra CL, Phillips ML, Quinlan ME. 2012. Autoinhibition of the formin Cappuccino in the absence of canonical autoinhibitory domains. Mol Biol Cell 23(19):3801–3813.

Bosch M, Le KH, Bugyi B, Correia JJ, Renault L, Carlier MF. 2007. Analysis of the function of Spire in actin assembly and its synergy with formin and profilin. Mol Cell 28(4):555–568.

Campellone KG, Welch MD. 2010. A nucleator arms race: cellular control of actin assembly. Nat Rev Mol Cell Biol 11(4):237–251.

Carlier MF, Jean C, Rieger KJ, Lenfant M, Pantaloni D. 1993. Modulation of the interaction between G-actin and thymosin beta 4 by the ATP/ADP ratio: possible implication in the regulation of actin dynamics. Proc Natl Acad Sci USA 90(11):5034–5038.

Carlier MF, Didry D, Erk I, Lepault J, Van Troys ML, Vandekerckhove J, Perelroizen I, Yin H, Doi Y, Pantaloni D. 1996. Tbeta 4 is not a simple G-actin sequestering protein and interacts with F-actin at high concentration. J Biol Chem 271(16):9231–9239.

Carlier MF, Husson C, Renault L, Didry D. 2011. Control of actin assembly by the WH2 domains and their multifunctional tandem repeats in Spire and Cordon-Bleu. Int Rev Cell Mol Biol 290:55–85.

Cheng HC, Skehan BM, Campellone KG, Leong JM, Rosen MK. 2008. Structural mechanism of WASP activation by the enterohaemorrhagic E. coli effector EspF(U). Nature 454(7207):1009–1013.

Chen Z, Borek D, Padrick SB, Gomez TS, Metlagel Z, Ismail AM, Umetani J, Billadeau DD, Otwinowski Z, Rosen MK. 2010. Structure and control of the actin regulatory WAVE complex. Nature 468(7323):533–538.

Chereau D, Kerff F, Graceffa P, Grabarek Z, Langsetmo K, Dominguez R. 2005. Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. Proc Natl Acad Sci USA 102(46):16644–16649.

Chhabra ES, Higgs HN. 2006. INF2 Is a WASP homology 2 motif-containing formin that severs actin filaments and accelerates both polymerization and depolymerization. J Biol Chem 281(36):26754–26767.

Chhabra ES, Ramabhadran V, Gerber SA, Higgs HN. 2009. INF2 is an endoplasmic reticulum-associated formin protein. J Cell Sci 122(Pt 9):1430–1440.

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325(5942):834–840.

Courtemanche N, Pollard TD. 2012. Determinants of Formin Homology 1 (FH1) domain function in actin filament elongation by formins. J Biol Chem 287(10):7812–7820.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14(6):1188–1190.

Dahlgaard K, Raposo AA, Niccoli T, St Johnston D. 2007. Capu and Spire assemble a cytoplasmic actin mesh that maintains microtubule organization in the Drosophila oocyte. Dev Cell 13(4):539–553.

De La Cruz EM, Ostap EM, Brundage RA, Reddy KS, Sweeney HL, Safer D. 2000. Thymosin-beta(4) changes the conformation and dynamics of actin monomers. Biophys J 78(5):2516–2527.

Didry D, Cantrelle FX, Husson C, Roblin P, Moorthy AM, Perez J, Le Clainche C, Hertzog M, Guittet E, Carlier MF and others. 2012. How a single residue in individual beta-thymosin/WH2 domains controls their functions in actin assembly. EMBO J 31(4):1000–1013.

Domanski M, Hertzog M, Coutant J, Gutsche-Perelroizen I, Bontems F, Carlier MF, Guittet E, van Heijenoort C. 2004. Coupling of folding and binding of thymosin beta4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. J Biol Chem 279(22):23637–23645.

Dominguez R. 2009. Actin filament nucleation and elongation factors–structure-function relationships. Crit Rev Biochem Mol Biol 44(6):351–366.

Ducka AM, Joel P, Popowicz GM, Trybus KM, Schleicher M, Noegel AA, Huber R, Holak TA, Sitar T. 2010. Structures of actin-bound Wiskott-Aldrich syndrome protein homology 2 (WH2) domains of Spire and the implication for filament nucleation. Proc Natl Acad Sci USA 107(26):11757–11762.

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6(3):197–208.

Eadie JS, Kim SW, Allen PG, Hutchinson LM, Kantor JD, Zetter BR. 2000. C-terminal variations in beta-thymosin family members specify functional differences in actin-binding properties. J Cell Biochem 77(2):277–287.

Egile C, Loisel TP, Laurent V, Li R, Pantaloni D, Sansonetti PJ, Carlier MF. 1999. Activation of the CDC42 effector N-WASP by the Shigella flexneri IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. J Cell Biol 146(6):1319–1332.

Fabrega C, Shen V, Shuman S, Lima CD. 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell 11(6):1549–1561.

Fuxreiter M. 2012. Fuzziness: linking regulation to protein dynamics. Mol Biosyst 8(1):168–177.

Fuxreiter M, Tompa P. 2012. Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol 725:1–14.

Gaucher JF, Mauge C, Didry D, Guichard B, Renault L, Carlier MF. 2012. Interactions of isolated C-terminal fragments of neural Wiskott-Aldrich syndrome protein (N-WASP) with actin and Arp2/3 complex. J Biol Chem 287(41):34646–34659.

Goode BL, Eck MJ. 2007. Mechanism and function of formins in the control of actin assembly. Annu Rev Biochem 76:593–627.

Gould CJ, Maiti S, Michelot A, Graziano BR, Blanchoin L, Goode BL. 2011. The formin DAD domain plays dual roles in autoinhibition and actin nucleation. Curr Biol 21(5):384–390.

Guenin-Mace L, Veyron-Churlet R, Thoulouze MI, Romet-Lemonne G, Hong H, Leadlay PF, Danckaert A, Ruf MT, Mostowy S, Zurzolo C and others. 2013. Mycolactone activation of Wiskott-Aldrich syndrome proteins underpins Buruli ulcer formation. J Clin Invest 123(4):1501–1512.

Guharoy M, Szabo B, Martos SC, Kosol S, Tompa P. 2013. Intrinsic structural disorder in cytoskeletal proteins. Cytoskeleton (Hoboken) in press; DOI: 10.1002/cm.21118.

Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. 2003. Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 28(2):81–85.

Hannappel E. 2010. Thymosin beta4 and its posttranslational modifications. Ann NY Acad Sci 1194:27–35.

Hansen MD, Kwiatkowski AV. 2013. Control of actin dynamics by allosteric regulation of actin binding proteins. Int Rev Cell Mol Biol 303:1–25.

Heimsath EG, Jr., Higgs HN. 2012. The C terminus of formin FMNL3 accelerates actin polymerization and contains a WH2 domain-like sequence that binds both monomers and filament barbed ends. J Biol Chem 287(5):3087–3098.

Hemsath L, Dvorsky R, Fiegen D, Carlier MF, Ahmadian MR. 2005. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. Mol Cell 20(2):313–324.

Hernandez-Valladares M, Kim T, Kannan B, Tung A, Aguda AH, Larsson M, Cooper JA, Robinson RC. 2010. Structural

characterization of a capping protein interaction motif defines a family of actin filament regulators. Nat Struct Mol Biol 17(4):497–503.

Hertzog M, van Heijenoort C, Didry D, Gaudier M, Coutant J, Gigant B, Didelot G, Preat T, Knossow M, Guittet E et al. 2004. The beta-thymosin/WH2 domain; structural basis for the switch from inhibition to promotion of actin assembly. Cell 117(5):611–623.

Hertzog M, Yarmola EG, Didry D, Bubb MR, Carlier MF. 2002. Control of actin dynamics by proteins made of beta-thymosin repeats: the actobindin family. J Biol Chem 277(17):14786–14792.

Hertzog M, Milanesi F, Hazelwood L, Disanza A, Liu H, Perlade E, Malabarba MG, Pasqualato S, Maiolica A, Confalonieri S, et al. 2010. Molecular basis for the dual function of Eps8 on actin dynamics: bundling and capping. PLoS Biol 8(6):e1000387.

Higgs HN, Pollard TD. 1999. Regulation of actin polymerization by Arp2/3 complex and WASp/Scar proteins. J Biol Chem 274(46): 32531–32534.

Huang H, Li L, Wu C, Schibli D, Colwill K, Ma S, Li C, Roy P, Ho K, Songyang Z, et al. 2008. Defining the specificity space of the human SRC homology 2 domain. Mol Cell Proteomics 7(4): 768–784.

Huff T, Muller CS, Otto AM, Netzker R, Hannappel E. 2001. beta-Thymosins, small acidic peptides with multiple functions. Int J Biochem Cell Biol 33(3):205–220.

Husson C, Cantrelle FX, Roblin P, Didry D, Le KH, Perez J, Guittet E, Van Heijenoort C, Renault L, Carlier MF. 2010. Multifunctionality of the beta-thymosin/WH2 module: G-actin sequestration, actin filament growth, nucleation, and severing. Ann NY Acad Sci 1194:44–52.

Husson C, Renault L, Didry D, Pantaloni D, Carlier MF. 2011. Cordon-Bleu uses WH2 domains as multifunctional dynamizers of actin filament assembly. Mol Cell 43(3):464–477.

Irobi E, Aguda AH, Larsson M, Guerin C, Yin HL, Burtnick LD, Blanchoin L, Robinson RC. 2004. Structural basis of actin sequestration by thymosin-beta4: implications for WH2 proteins. EMBO J 23(18):3599–3608.

Kelly AE, Kranitz H, Dotsch V, Mullins RD. 2006. Actin binding to the central domain of WASP/Scar proteins plays a critical role in the activation of the Arp2/3 complex. J Biol Chem 281(15):10589–10597.

Kim AS, Kakalis LT, Abdul-Manan N, Liu GA, Rosen MK. 2000. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. Nature 404(6774):151–158.

Kim T, Cooper JA, Sept D. 2010. The interaction of capping protein with the barbed end of the actin filament. J Mol Biol 404(5): 794–802.

Kim T, Ravilious GE, Sept D, Cooper JA. 2012. Mechanism for CARMIL protein inhibition of heterodimeric actin-capping protein. J Biol Chem 287(19):15251–15262.

Koronakis V, Hume PJ, Humphreys D, Liu T, Horning O, Jensen ON, McGhie EJ. 2011. WAVE regulatory complex activation by cooperating GTPases Arf and Rac1. Proc Natl Acad Sci USA 108(35):14449–14454.

Lammers M, Rose R, Scrima A, Wittinghofer A. 2005. The regulation of mDia1 by autoinhibition and its release by Rho*GTP. EMBO J 24(23):4176–4187.

Lee SH, Kerff F, Chereau D, Ferron F, Klug A, Dominguez R. 2007. Structural basis for the actin-binding function of missing-in-metastasis. Structure 15(2):145–155.

Letunic I, Doerks T. Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40: D302–D305.

Li F, Higgs HN. 2005. Dissecting requirements for auto-inhibition of actin nucleation by the formin, mDia1. J Biol Chem 280(8): 6986–6992.

Liverman AD, Cheng HC, Trosky JE, Leung DW, Yarbrough ML, Burdette DL, Rosen MK, Orth K. 2007. Arp2/3-independent assembly of actin by Vibrio type III effector VopL. Proc Natl Acad Sci USA 104(43):17117–17122.

Manseau LJ, Schupbach T. 1989. cappuccino and spire: two unique maternal-effect loci required for both the anteroposterior and dorsoventral patterns of the Drosophila embryo. Genes Dev 3(9):1437–1452.

Marchand JB, Kaiser DA, Pollard TD, Higgs HN. 2001. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. Nat Cell Biol 3(1):76–82.

Mattila PK, Salminen M, Yamashiro T, Lappalainen P. 2003. Mouse MIM, a tissue-specific regulator of cytoskeletal dynamics, interacts with ATP-actin monomers through its C-terminal WH2 domain. J Biol Chem 278(10):8452–8459.

McNulty BC, Young GB, Pielak GJ. 2006. Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. J Mol Biol 355(5):893–897.

Mendoza MC. 2013. Phosphoregulation of the WAVE regulatory complex and signal integration. Semin Cell Dev Biol 24(4):272–279.

Mouilleron S, Guettler S, Langer CA, Treisman R, McDonald NQ. 2008. Molecular basis for G-actin binding to RPEL motifs from the serum response factor coactivator MAL. EMBO J 27(23):3198–3208.

Mouilleron S, Wiezlak M, O'Reilly N, Treisman R, McDonald NQ. 2012. Structures of the Phactr1 RPEL domain and RPEL motif complexes with G-actin reveal the molecular basis for actin binding cooperativity. Structure 20(11):1960–1970.

Narita A, Takeda S, Yamashita A, Maeda Y. 2006. Structural basis of actin filament capping at the barbed-end: a cryo-electron microscopy study. EMBO J 25(23):5626–5633.

Oikawa T, Yamaguchi H, Itoh T, Kato M, Ijuin T, Yamazaki D, Suetsugu S, Takenawa T. 2004. PtdIns(3,4,5)P3 binding is necessary for WAVE2-induced formation of lamellipodia. Nat Cell Biol 6(5):420–426.

Otomo T, Tomchick DR, Otomo C, Panchal SC, Machius M, Rosen MK. 2005. Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain. Nature 433(7025):488–494.

Padrick SB, Rosen MK. 2010. Physical mechanisms of signal integration by WASP family proteins. Annu Rev Biochem 79:707–735.

Panchal SC, Kaiser DA, Torres E, Pollard TD, Rosen MK. 2003. A conserved amphipathic helix in WASP/Scar proteins is essential for activation of Arp2/3 complex. Nat Struct Biol 10(8):591–598.

Paul AS, Pollard TD. 2008. The role of the FH1 domain and profilin in formin-mediated actin-filament elongation and nucleation. Curr Biol 18(1):9–19.

Paul AS, Pollard TD. 2009. Review of the mechanism of processive actin filament elongation by formins. Cell Motil Cytoskeleton 66(8):606–617.

Paunola E, Mattila PK, Lappalainen P. 2002. WH2 domain: a small, versatile adapter for actin monomers. FEBS Lett 513(1):92–97.

Perelroizen I, Marchand JB, Blanchoin L, Didry D, Carlier MF. 1994. Interaction of profilin with G-actin and poly(L-proline). Biochemistry 33(28):8472–8478.

Peterson JR, Bickford LC, Morgan D, Kim AS, Ouerfelli O, Kirschner MW, Rosen MK. 2004. Chemical inhibition of N-WASP

by stabilization of a native autoinhibited conformation. Nat Struct Mol Biol 11(8):747–755.

Petrella EC, Machesky LM, Kaiser DA, Pollard TD. 1996. Structural requirements and thermodynamics of the interaction of proline peptides with profilin. Biochemistry 35(51):16535–16543.

Pfender S, Kuznetsov V, Pleiser S, Kerkhoff E, Schuh M. 2011. Spire-type actin nucleators cooperate with Formin-2 to drive asymmetric oocyte division. Curr Biol 21(11):955–960.

Pollard TD, Cooper JA. 2009. Actin, a central player in cell shape and movement. Science 326(5957):1208–1212.

Pontius BW. 1993. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends Biochem Sci 18(5):181–186.

Qualmann B, Kessels MM. 2009. New players in actin polymerization–WH2-domain-containing actin nucleators. Trends Cell Biol 19(6):276–285.

Quinlan ME, Heuser JE, Kerkhoff E, Mullins RD. 2005. Drosophila Spire is an actin nucleation factor. Nature 433(7024):382–388.

Quinlan ME, Hilgert S, Bedrossian A, Mullins RD, Kerkhoff E. 2007. Regulatory interactions between two actin nucleators, Spire and Cappuccino. J Cell Biol 179(1):117–128.

Rebowski G, Namgoong S, Boczkowska M, Leavis PC, Navaza J, Dominguez R. 2010. Structure of a longitudinal actin dimer assembled by tandem w domains: implications for actin filament nucleation. J Mol Biol 403(1):11–23.

Renault L, Bugyi B, Carlier MF. 2008. Spire and Cordon-bleu: multifunctional regulators of actin dynamics. Trends Cell Biol 18(10):494–504.

Rottner K, Hanisch J, Campellone KG. 2010. WASH, WHAMM and JMY: regulation of Arp2/3 complex and beyond. Trends Cell Biol 20(11):650–661.

Safer D, Sosnick TR, Elzinga M. 1997. Thymosin beta 4 binds actin in an extended conformation and contacts both the barbed and pointed ends. Biochemistry 36(19):5806–5816.

Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 97(16):8868–8873.

Simenel C, Van Troys M, Vandekerckhove J, Ampe C, Delepierre M. 2000. Structural requirements for thymosin beta4 in its contact with actin. An NMR-analysis of thymosin beta4 mutants in solution and correlation with their biological activity. Eur J Biochem 267(12):3530–3538.

Sosne G, Qiu P, Goldstein AL, Wheater M. 2010. Biological activities of thymosin {beta}4 defined by active sites in short peptide sequences. FASEB J 24(7):2144–2151.

Sribenja S, Wongkham S, Wongkham C, Yao Q, Chen C. 2013. Roles and mechanisms of beta-thymosins in cell migration and cancer metastasis: an update. Cancer Invest 31(2):103–110.

Staus DP, Taylor JM, Mack CP. 2011. Enhancement of mDia2 activity by Rho-kinase-dependent phosphorylation of the diaphanous autoregulatory domain. Biochem J 439(1):57–65.

Stoll R, Voelter W, Holak TA. 1997. Conformation of thymosin beta 9 in water/fluoroalcohol solution determined by NMR spectroscopy. Biopolymers 41(6):623–634.

Takeda S, Minakata S, Koike R, Kawahata I, Narita A, Kitazawa M, Ota M, Yamakuni T, Maeda Y, Nitanai Y. 2010. Two distinct mechanisms for actin capping protein regulation–steric and allosteric inhibition. PLoS Biol 8(7):e1000416.

Takeda S, Koike R, Nitanai Y, Minakata S, Maeda Y, Ota M. 2011. Actin capping protein and its inhibitor CARMIL: how intrinsically disordered regions function. Phys Biol 8(3):035005.

Takeya R, Taniguchi K, Narumiya S, Sumimoto H. 2008. The mammalian formin FHOD1 is activated through phosphorylation by ROCK and mediates thrombin-induced stress fibre formation in endothelial cells. EMBO J 27(4):618–628.

Tam VC, Serruto D, Dziejman M, Brieher W, Mekalanos JJ. 2007. A type III secretion system in Vibrio cholerae translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization. Cell Host Microbe 1(2):95–107.

Thompson ME, Heimsath EG, Gauvin TJ, Higgs HN, Kull FJ. 2013. FMNL3 FH2-actin structure gives insight into formin-mediated actin nucleation and elongation. Nat Struct Mol Biol 20(1):111–118.

Tompa P. 2003. Intrinsically unstructured proteins evolve by repeat expansion. Bioessays 25(9):847–855.

Tompa P. 2011. Unstructural biology coming of age. Curr Opin Struct Biol 21(3):419–425.

Tompa P. 2012. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37(12):509–516.

Tompa P, Szasz C, Buday L. 2005. Structural disorder throws new light on moonlighting. Trends Biochem Sci 30(9):484–489.

Uversky VN. 2010. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. Expert Rev Proteomics 7(4):543–564.

Uversky VN, Dunker AK. 2010. Understanding protein non-folding. Biochim Biophys Acta 1804(6):1231–1264.

Van Troys M, Ono K, Dewitte D, Jonckheere V, De Ruyck N, Vandekerckhove J, Ono S, Ampe C. 2004. TetraThymosinbeta is required for actin dynamics in Caenorhabditis elegans and acts via functionally different actin-binding repeats. Mol Biol Cell 15(10):4735–4748.

Vizcarra CL, Kreutz B, Rodal AA, Toms AV, Lu J, Zheng W, Quinlan ME, Eck MJ. 2011. Structure and function of the interacting domains of Spire and Fmn-family formins. Proc Natl Acad Sci USA 108(29):11884–11889.

Wallar BJ, Stropich BN, Schoenherr JA, Holman HA, Kitchen SM, Alberts AS. 2006. The basic region of the diaphanous-autoregulatory domain (DAD) is required for autoregulatory interactions with the diaphanous-related formin inhibitory domain. J Biol Chem 281(7):4300–4307.

Winder SJ, Ayscough KR. 2005. Actin-binding proteins. J Cell Sci 118(Pt 4):651–654.

Yamaguchi H, Miki H, Suetsugu S, Ma L, Kirschner MW, Takenawa T. 2000. Two tandem verprolin homology domains are necessary for a strong activation of Arp2/3 complex-induced actin polymerization and induction of microspike formation by N-WASP. Proc Natl Acad Sci USA 97(23):12631–12636.

Zeth K, Pechlivanis M, Samol A, Pleiser S, Vonrhein C, Kerkhoff E. 2011. Molecular basis of actin nucleation factor cooperativity: crystal structure of the Spir-1 kinase non-catalytic C-lobe domain (KIND)*formin-2 formin SPIR interaction motif (FSI) complex. J Biol Chem 286(35):30732–30739.

Zuchero JB, Coutts AS, Quinlan ME, Thangue NB, Mullins RD. 2009. p53-cofactor JMY is a multifunctional actin nucleation factor. Nat Cell Biol 11(4):451–459.

Zuchero JB, Belin B, Mullins RD. 2012. Actin binding to WH2 domains regulates nuclear import of the multifunctional actin regulator JMY. Mol Biol Cell 23(5):853–863.

Zwolak A, Fujiwara I, Hammer JA, 3rd, Tjandra N. 2010. Structural basis for capping protein sequestration by myotrophin (V-1). J Biol Chem 285(33):25767–25781.

Journal of Structural Biology 188 (2014) 71–78

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Robust and low cost uniform ^{15}N-labeling of proteins expressed in *Drosophila* S2 cells and *Spodoptera frugiperda* Sf9 cells for NMR applications

Annalisa Meola [a,b,c,1], Célia Deville [a,1], Scott A. Jeffers [b,c], Pablo Guardado-Calvo [b,c], Ieva Vasiliauskaite [b,c], Christina Sizun [a], Christine Girard-Blanc [d], Christian Malosse [e,f], Carine van Heijenoort [a], Julia Chamot-Rooke [e,f], Thomas Krey [b,c], Eric Guittet [a], Stéphane Pêtres [d], Félix A. Rey [b,c], François Bontems [a,b,c,*]

[a] *Laboratoire de chimie et biologie structurales, Institut de chimie des substances naturelles, CNRS UPR2301, Centre de recherche de Gif-sur-Yvette, 91190 Gif-sur-Yvette, France*
[b] *Département de virologie, Unité de virologie structurale, Institut Pasteur, 75015 Paris, France*
[c] *CNRS URA3015, 75015 Paris, France*
[d] *Plateforme de production de protéines recombinantes, Institut Pasteur, 75015 Paris, France*
[e] *Unité de spectrométrie de masse structurale et protéomique, Institut Pasteur, 75015 Paris, France*
[f] *CNRS UMR 3528, Institut Pasteur, 75015 Paris, France*

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form 15 June 2014
Accepted 15 August 2014
Available online 28 August 2014

Keywords:
NMR structural studies of eukaryotic proteins
^{15}N-labeling of proteins produced in Sf9 and S2 insect cells
Cellular and viral membrane fusion proteins
Non-polymerizable actin mutant
Viral glycoprotein domain characterization

ABSTRACT

Nuclear magnetic resonance spectroscopy is a powerful tool to study structural and functional properties of proteins, provided that they can be enriched in stable isotopes such as ^{15}N, ^{13}C and ^{2}H. This is usually easy and inexpensive when the proteins are expressed in *Escherichia coli*, but many eukaryotic (human in particular) proteins cannot be produced this way. An alternative is to express them in insect cells. Labeled insect cell growth media are commercially available but at prohibitive prices, limiting the NMR studies to only a subset of biologically important proteins. Non-commercial solutions from academic institutions have been proposed, but none of them is really satisfying. We have developed a ^{15}N-labeling procedure based on the use of a commercial medium depleted of all amino acids and supplemented with a ^{15}N-labeled yeast autolysate for a total cost about five times lower than that of the currently available solutions. We have applied our procedure to the production of a non-polymerizable mutant of actin in Sf9 cells and of fragments of eukaryotic and viral membrane fusion proteins in S2 cells, which typically cannot be produced in *E. coli*, with production yields comparable to those obtained with standard commercial media. Our results support, in particular, the putative limits of a self-folding domain within a viral glycoprotein of unknown structure.

1. Introduction

Nuclear magnetic resonance spectroscopy is a powerful and versatile technique to study the structural and functional properties of proteins. Modern biological NMR applications rely on stable isotope labeling, from uniform enrichment of the molecule with ^{15}N and ^{13}C to specific introduction of ^{1}H,^{13}C-labeled methyl groups in a ^{2}H,^{12}C background. These labeling schemes are generally straightforward when the protein can be over-expressed in *Escherichia coli*. This bacterium can be grown using minimal medium with NH_4Cl and glucose or glycerol as nitrogen and carbon sources, respectively. Moreover *E. coli* tolerates the replacement of H_2O by $^{2}H_2O$ for the production of deuterated samples. Labeled amino acids or amino acid precursors can be added to the medium to obtain specifically labeled samples. An important aspect is that these techniques are usually not only easy but also inexpensive. For many proteins, it is possible to obtain at least one NMR sample from one liter of culture containing 1 g of $^{15}NH_4Cl$ (~30 euros), 2–3 g of ^{13}C-glucose (~300 euros) and in the case of deuterated samples, 1 l of $^{2}H_2O$ (~350 euros). However, many eukaryotic (human in particular) proteins cannot be produced in *E. coli* because they need specific chaperones and/or post-translational modifications to be correctly folded and functional. Although it is sometimes

* Corresponding author at: Laboratoire de chimie et biologie structurales, Institut de chimie des substances naturelles, CNRS UPR2301, Centre de recherche de Gif-sur-Yvette, 91190 Gif-sur-Yvette, France.
E-mail address: francois.bontems@cnrs.fr (F. Bontems).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.jsb.2014.08.002
1047-8477/

feasible to over-express them in yeast, in most cases it is necessary to use cells from higher organisms. The most common are insect cells, easier to work with than mammalian cells and able to provide the relevant post-translational modifications. Insect cells are heterotrophic organisms only able to grow on complex media, the most efficient for protein production being proprietary media of unreleased composition. Some of them are sold for NMR applications but at prohibitive prices (about 10,000 euros per liter for *Cambridge Isotope Laboratories* ^{15}N,^{13}C-BioExpress 2000), *de facto* precluding the production of labeled samples for most NMR laboratories and strongly restricting the biologically relevant eukaryotic systems that can be studied by this technique.

This leads to a paradoxical situation. On the one hand, there is a strong interest in the NMR community for labeling proteins expressed in eukaryotic cells (Takashi and Shimada, 2010; Saxena et al., 2012; Gossert and Jahnke, 2012). But on the other hand, labeling in eukaryotic cells has remained confined to only a few examples, including amino acid type selective labeling, which can lead to interesting applications but is limited to very specific cases (Werner et al., 2008; Nygaard et al., 2013). Uniform labeling was carried out, using the aforementioned commercial media, for the study of the Abelson kinase domain (Strauss et al., 2005) and HIV gp120 glycoprotein (Sastry et al., 2011), and for the observation of proteins *in cellula* (Hamatsu et al., 2013). Several laboratories have proposed self-made uniform-labeling media as an alternative to the commercial one. The Fesik (Hansen et al., 1992) and Logan (Walton et al., 2006) groups respectively supplemented classical media with ^{15}N-algae extract and ^{15}N-Cys, ^{15}N-Trp, ^{15}N-Gln (Hansen et al., 1992) or ^{15}N-Cys, ^{15}N-Trp, ^{15}N-Gln, ^{15}N-Glu, $^{15}NH_4Cl$ (Walton et al., 2006), but both solutions remained expensive due to the use of ^{15}N-labeled amino acids and led to reduced production yields as compared to commercial media. In 2011, DeGrip and colleagues proposed a medium formulation based on the DMEM/F12 recipe for CHO and HEK293 mammalian cells supplemented with a homemade labeled yeast autolysate and a small amount of algae hydrolysate. They obtained protein yields, compared to standard DMEM, that were about 40–60% in CHO and 60–80% in HEK293 cells (Egorova-Zachernyuk et al., 2011) One of the authors created a startup company (*Protein labeling innovation*) to sell yeast autolysates at about 1400 and 4000 euros per gram for ^{15}N and ^{15}N,^{13}C-labeling, respectively. Their media are slightly less expensive than those of *Cambridge Isotope Laboratories*, but there is no information concerning the resulting production efficiency as compared to the use of optimized commercial media.

Here, we demonstrated that it is feasible to prepare a simple, low cost and efficient medium for the production of uniformly ^{15}N-labeled proteins in both *Drosophila* Schneider 2 (S2) and *Spodoptera frugiperda* Sf9 cells. This particular medium, which gives similar protein yields as standard commercial ones, is derived from a commercial medium depleted in amino acids (ESF921-Δaa from *Expression Systems* in our case, but other amino acids depleted media are also available) supplemented with 10 g l^{-1} of a ^{15}N-labeled yeast autolysate designed *ad-hoc* by the *Bio Springer* company (now distributed by *Cortecnet*) and 0.27 g l^{-1} (5 mM) $^{15}NH_4Cl$ as glutamine replacement. We also show that it is possible to transfer both cells types from a standard medium to the new labeling medium without affecting the yield, as already proposed for the Sf9 cells but between two commercially optimized media (Strauss et al., 2005; Hamatsu et al., 2013). We further applied this approach to two biological systems requiring specific chaperones and eukaryotic post-translational modifications: a non-polymerizable mutant of *Drosophila* actin (Rould et al., 2006) in Sf9 cells, (to study its interaction with unstructured peptides regulating its polymerization), and an autonomously folded domain of both cellular and viral class II membrane fusion proteins in S2 cells.

2. Experimental section

2.1. Materials

^{15}N-labeled dried yeast autolysate (YA) was provided by *Bio Springer* and is now distributed by *Cortecnet* (*ref*: *YA15N*). Regular and amino acid depleted ESF921 (*ref: 96-275 921 AA deficient, available upon request*) insect cell culture media were purchased from *Expression Systems*. Insect-Xpress insect cell culture medium was purchased from *Lonza*. The yeast autolysate containing media were prepared by dissolving the dry yeast powder in the amino acid depleted ESF921 medium at final concentrations of 20 (YA_{20}-ESF), 10 (YA_{10}-ESF) or 5 g l^{-1} (YA_5-ESF), by adding either 1% of 200 mM L-glutamine solution (*Gibco*) or 0.27 g l^{-1} (5 mM) of $^{15}NH_4Cl$ and by sterilizing the solution using 0.22 μm filters (*Millipore*).

2.2. Cell culture and protein expression

Classically, S2 cells are amplified in tissue culture flasks until they reach a density of $\sim 10^7$ cells ml^{-1} in a volume of ~200 ml. The content of the flasks is then transferred to a spinner that is gradually filled as the number of cells increases. Protein production is induced once the spinner is full. In the case of Sf9/baculovirus, the cells are similarly amplified, centrifuged at the end of the amplification stage, resuspended in a small volume of virus solution and transferred to the spinner together with fresh medium. The advantage of the Sf9/baculovirus protocol is that cell amplification is carried out in a non-labeled (less expensive) medium before transfer into the labeling medium. For the same reason, we replaced the medium at the end of the amplification stage of the S2 cells. This also permits to increase the cell density beyond the values classically used and thereby to decrease the required volume of labeling medium. Accordingly, S2 cells protein production tests were done in T-150 tissue culture flasks (*Corning Incorporated*) at final volumes between 15 ml (one flask) and 45 ml (three flasks) and densities of $\sim 30.10^6$ cells ml^{-1}. The cells were successively amplified in T-25, T-75 and T-150 flasks until the desired number of cells was obtained. They were gently centrifuged (5 min at 100 *g*) and resuspended in the production medium. Protein expression was induced with 4 μM $CdCl_2$ just after transfer into the production medium. Cells were removed 7 days after.

Expression of proteins in Sf9 cell was performed in 500 ml spinners containing 300–500 ml of expression medium. Typically, $1500{\cdot}10^6$ cells were centrifuged and mixed with $7500{\cdot}10^6$ baculoviruses (MOI of five). After one-hour incubation at room temperature, the cells were transferred to the spinner and the volume was adjusted with the expression medium supplemented with gentamicin (50 mg l^{-1}). Protein expression was carried out for 3 days before the cells were harvested and stored at −20 °C.

2.3. Proteins and protein purification

Five proteins or protein fragments were produced during this work: DIII-EFF-1 (Arg409-Leu609) and DIII-AFF-1 (Ser388-Ala485, Mw 15 kDa) designed to correspond to the domains III of the *Caenorhabditis elegans* EFF-1 and AFF-1 fusion proteins (Peres-Varga et al., in press; Podbilewicz et al., 2006; Sapir et al., 2007), DIII-HaV (Met283-Asp415, Mw 18 kDa) designed to comprise the domain III of the Hantaan virus Gc fusion protein based on the prediction for its ortholog in Andes hantavirus (Tischler et al., 2005), DIII-RVFV (Ala336-Ser457, Mw 17 kDa) corresponding to domain III of the Rift valley fever virus Gc protein (Dessau and

Modis, 2013) and finally AP-actin (Mw 42 kDa) corresponding to a non-polymerizable double mutant of actin (Rould et al., 2006). DIII-EFF-1, DIII-AFF-1, DIII-HaV and DIII-RVFV were fused with a signal peptide for efficient co-translational translocation into the endoplasmic reticulum of S2 cells. AP-actin was produced intracellularly in Sf9 cells with a recombinant baculovirus. All proteins were fused to a double strep-tag (N-terminally for actin, C-terminally for the others) and purified on a one-milliliter Strep-Trap column using an Äkta purifier system (*GE Healthcare*). Since aggregates drastically reduce actin sample lifetime, AP-actin was further purified on a Superdex 75 HiLoad 16/60 size exclusion column (*GE Healthcare*). Protein samples were stored and analyzed in PBS at pH 6.4 except for AP-actin, stored and analyzed in 50 mM Tris HCl pH 7.4, 0.5 mM $CaCl_2$, 0.5 mM ATP, 1 mM DTT. Final sample concentrations were estimated by measuring absorption at 280 nm.

2.4. NMR spectroscopy

Estimation of media amino acid content was performed on a Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe and a SampleJet sample changer, by comparing ^{1}H spectra of the media alone and with incrementally added quantities (0.25 and 0.50 g l^{-1}) of each amino acids. The other spectra were recorded on Bruker Avance III 800 and 950 MHz spectrometers equipped with TCI cryoprobes at 298 K. ^{15}N-HSQC and ^{15}N-sofast-HMQC spectra were acquired with 32 scans and 200 increments for 29 ppm in the ^{15}N dimension and 128 scans and 512 increments for 60 ppm, respectively. Isotopic incorporation was estimated by recording two-dimensional ^{1}H experiments reduced to their diagonal (equivalent to 2D-NOESY spectra with a null mixing time), ^{15}N decoupled during the non-physical *t*1 labeling period but not during the *t*2 acquisition times, leading to three lines for each frequency in the *t*2 dimension, the central line corresponding to ^{1}H bound to ^{14}N and the peripheral to ^{1}H bound to ^{15}N. Spectra were processed with Topspin 3.1 (Bruker) and analyzed with CCPNMR (Vranken et al., 2005).

2.5. Mass spectrometry

A simple cleaning step using C_4 Ziptip® was used for all samples and proteins of interest were eluted in 10 µl (MeOH/H_2O/HCOOH) (73.5/23.5/3) (v/v/v). Desalted proteins were further analyzed in positive ion mode by direct infusion nanoelectrospray, using a TriVersa Nanomate (*Advion Biosciences*) on a LTQ Velos Orbitrap mass spectrometer, equipped with ETD module (*Thermo Fisher Scientific*). A full set of automated positive ion calibrations was performed immediately prior to mass measurement. All spectra were acquired in full profile mode. For MS experiments ions were accumulated in the ion trap and then transferred to the Orbitrap for high resolution mass measurement. For MS/MS experiments, ions were selected with an appropriate mass window and HCD was performed at normalized collision energies of 15–25%, with other activation parameters left as default. The FT automatic gain control (AGC) was set at $1{\cdot}10^{6}$ for MS and $2{\cdot}10^{5}$ for MSn experiments. Spectra were acquired in the FTMS over several minutes with one microscan and a resolution of 60,000 for MS and 30,000 @ *m*/*z* 400 for MS/MS before being processed with Thermo Xcalibur2.2. Multiply charged ions spectra were deconvoluted using Xtract. Isotopic enrichment was calculated using the ratio between the experimental average mass of the labeled protein and the average mass of the same protein computed for complete heavy isotope incorporation. Accurate molecular mass and MS/MS were used to check the sequence of all proteins of interest.

3. Results and discussion

3.1. Media formulation

The total amount of amino acids in ESF921 is above 10 g l^{-1}. This value is similar to that provided by *BioConcept* for Sf4 medium (13 g l^{-1}). An evaluation by *Bio Springer* of the content of free amino acids in the yeast autolysate gives a value of 50% (w/w). Titration of amino acids in YA_{20}-ESF is in agreement with this value. The amino acid profiles of the YA_{20}-ESF and ESF921 media are rather different with an overrepresentation in YA_{20}-ESF of alanine, glycine, leucine, serine and threonine and an underrepresentation of methionine and glutamine. However, ESF921 and Sf4 are also different, showing that the cells are able to adapt to different amino acid compositions. Since DeGrip and collaborators (Egorova-Zachernyuk et al., 2009) obtained the best growth rate with 4 g l^{-1} of autolysate, with a net decrease at 8 g l^{-1}, we tested three media with yeast autolysate concentrations of 20, 10 and 5 g l^{-1}.

3.2. Glutamine replacement by ammonium chloride

Insect cell growth media contain large amounts of glutamine (above 6 mM) necessary for protein synthesis but also for energetic metabolism. Glutamine is strongly underrepresented in YA_{20}-ESF (about 1.2 mM). A solution could be to add ^{15}N-glutamine (Hansen et al., 1992; Walton et al., 2006) but at the expense of an important increase of the medium cost. Fesik and collaborators enzymatically synthesized ^{15}N-glutamine from ^{15}N-glutamate and $^{15}NH_4Cl$, but this solution remains expensive and imposes additional steps that we wanted to avoid. However, as shown by Häggstrom and collaborators, glutamine is not essential for Sf9 cells and can be replaced by 5 mM of NH_4Cl with the same growth performance (Öhman et al., 1995). In Table 1, we compare the amounts of protein produced by S2 cells with Insect-Xpress in spinner and with either YA_{20}-ESF supplemented with 2 mM glutamine or YA_{10}-ESF supplemented with 5 mM $^{15}NH_4Cl$. In the latter two cases, cells were first adapted to the labeling media and the production was done in flasks. The amount of protein obtained after strep-tag purification with respect to the culture volume is similar in the three conditions. It is thus possible to replace the glutamine by NH_4Cl and to lower the amount of yeast autolysate from 20 to 10 g l^{-1} without yield losses. There is no difference between the protein concentrations obtained in spinners with a density of $\sim 8.10^{6}$ cells ml^{-1} and in the flasks with densities of $\sim 30.10^{6}$ - cells ml^{-1}. A first hypothesis to explain this is that the limiting factor is not the cell quantity but an element present in the medium. A second is to note that the cells used for productions in spinners were only grown in Insect-Xpress while those used for flask expression were adapted from Insect-Xpress to ESF921 and from ESF921 to YA-ESF before amplification, resulting in higher stress for the cells, which become less productive.

3.3. Direct transfer of the S2 cells in the labeling medium

Cell adaptation in YA-ESF medium before amplification is very constraining, requiring two steps from Insect-Xpress (standard

Table 1
Comparison of DIII-EFF production yields obtained in S2 cells in spinners with Insect-Xpress and in flasks with YA_{20}-ESF and glutamine or YA_{10}-ESF and 5 mM $^{15}NH_4Cl$.

DIII-EFF	mg l^{-1}	nmol ml^{-1}	pmol Mc^{-1}
YA_{20}-ESF + Q[a]	6.0	0.4	14
YA_{10}-ESF + NH_4Cl[a]	10.5	0.7	23
IXP spinner[b]	~9	~0.6	~75

[a] Induced at 30.10^{6} cells ml^{-1}.
[b] Induced at 8.10^{6} cells ml^{-1}.

expression medium in the laboratory) to ESF921 and from ESF921 to YA-ESF. This adds 2–3 weeks to the production time. Moreover, the viability of cells grown in Insect-XPress, ESF921 and YA-ESF was stable during several weeks, but its evolution in YA-ESF in the long term was not known. In their in-cell NMR analysis, Hamatsu et al. (2013) directly transferred their cells from unlabeled sf900II (*GIBCO*) to labeled Bioexpress 2000 (*CIL*). However, the transfer was done between two amino-acid rich media and the cells only needed to survive for a few hours in the spectrometer. In our case, the labeling medium was much less rich in amino acids and we wanted the S2 cells to produce the protein during at least one week. We compared the survival of S2 cells expressing two different proteins grown in ESF921 and transferred in ESF921, YA_{10}-ESF or YA_5-ESF just before the induction (Fig. 1). One week after induction there were no significant differences in the percentage of living cells (between 41% and 48%) in the three media. In all experiments the number of living cells remained globally constant, likely because cell mortality is compensated with cell division. This indicates that it is possible to remove at least the second adaptation phase (from ESF921 to YA-ESF), but that it is also possible to use 5 g l^{-1} of yeast autolysate without compromising cell viability.

3.4. Production efficiency of S2 and Sf9 cells

As the direct transfer of the cells from ESF921 into YA-ESF media is well tolerated by S2 cells, we simplified our protocol by growing both S2 and Sf9 cells in Insect-Xpress and transferring them into YA-ESF just before induction. In addition, according to our preliminary experiments and to the literature (Egorova-Zachernyuk et al., 2009), we tried both 10 and 5 g l^{-1} of yeast autolysate in the YA-ESF media. The protein production yields obtained on different systems under these conditions are summarized in Table 2. In all cases, YA_{10}-ESF gives results comparable to

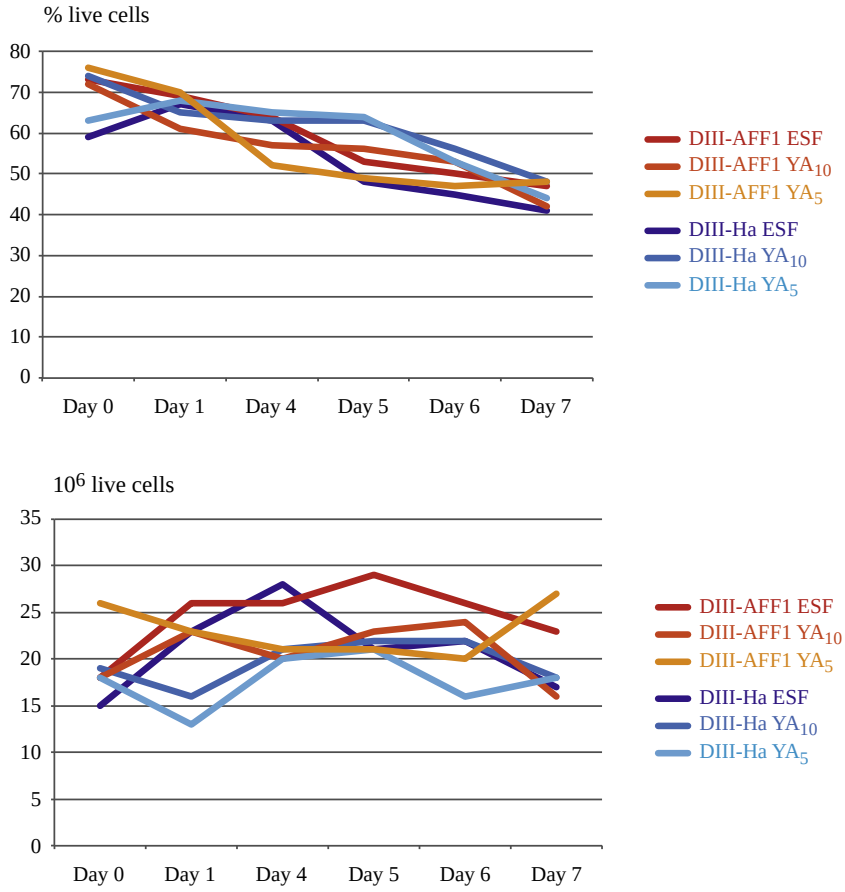


Fig.1. Evolution of the percentage (upper panel) or the number (lower panel) of living S2 cells expressing DIII-HaV or DIII-AFF-1 after induction in ESF921, YA_{10}-ESF or YA_5-ESF media. In all cases, the cells were amplified in ESF921, centrifuged and resuspended in the expression media just prior the induction. The graphs clearly show that the three media are comparable in terms of cell survival.

Table 2
Comparison of protein expression yields obtained with S2 and Sf9 cells with Insect-Xpress in spinner or flasks and with YA_{10}-ESF or YA_5-ESF in flasks without adaptation.

	Spinner[a]	IXP[b]	YA_{10}-ESF[b]	YA_5-ESF[b]	ml for 20 nmol[c]
S2, DIII-HaV	15	22 [18–28]	17 [13–19]	11 [16–7]	22
S2, DIII-RVFV	8	10	12		28
S2, DIII-AFF1	4	4	4		75
Sf9, actin	6		10 [9–10]	6	83

[a] Induced at 8.10^6 cells ml^{-1}. The values (in mg l^{-1}) correspond to those classically obtained in the *Unité de virologie structurale* for structural applications.
[b] Cells were amplified in Insect-Xpress and transferred in the expression medium without adaptation, at a density of 30.10^6 cells ml^{-1}. In the case of DIII-HaV and of AP-actin, productions were realized over a period of several months, the mean, minimal and maximal values (measured after strep-tag purification) are indicated in mg l^{-1}.
[c] Volume necessary to produce about 20 nmol of the proteins, this value corresponding to an NMR sample of 100 μM in 200 μl (sample volume in a 3 mm NMR tube).

those obtained with Insect-Xpress. For both S2 and Sf9 cells, the amount of protein obtained with the new medium is slightly higher than what is usually obtained in the laboratory with Insect-Xpress in spinners. The yields are similar (from 70% to 120%) when the S2 cells are centrifuged from Insect-Xpress and resuspended at high concentration in either Insect-Xpress or YA_{10}-ESF. In the case of the actin mutant, we verified that the amount of aggregated protein at the end of the production process was the same in both media (data not shown). This shows that both S2 and Sf9 cells tolerate the direct transfer from Insect-Xpress to YA_{10}-ESF without production losses, despite lower amino acid content. Comparison between the yields obtained with Insect-Xpress in spinners, at low concentration without replacement of the medium, and in flasks, at high concentration with medium replacement, indicates that the latter procedure increases the production yield in the case of the S2 cells. Importantly, we performed these experiments by using two different yeast autolysate batches (one treated to remove protease activity, see below) resulting from two different yeast cultures without noticing differences in productions yields. In the case of YA_5-ESF, the yields are generally lower than those obtained in Insect-Xpress. The yields in S2 cells are sometimes comparable to those obtained in spinners, but the variability is much larger.

3.5. Protease activity inhibition

In the first production assays of the DIII-HaV fragment, a heterogeneous product was obtained in Insect-Xpress as well as in YA-ESF. In Insect-Xpress, a protein with the expected mass (18 kDa) was purified but shorter forms were also observed (Fig. 2A). In YA-ESF, only the shortest form (~15 kDa) was recovered. While in Insect-Xpress the protein was subject to partial degradation likely due to proteases released by the cells, degradation was strongly enhanced in YA-ESF. Since we suspected the presence of proteases in the yeast autolysate, *Bio Springer* modified the yeast autolysate production protocol by including a protease inactivation step. Comparison of the product obtained after incubating the purified protein in the different media indeed showed complete degradation in the initial (non-treated) YA-ESF but no degradation in Insect-Xpress and in the treated YA-ESF (Fig. 2B). Subsequent production assays confirmed the absence of degradation in treated YA-ESF (Fig. 2C) and showed that the inactivation step had no influence on the production yields.

3.6. ^{15}N incorporation

^{15}N incorporation was evaluated either by mass spectrometry or by NMR spectroscopy (Fig. 3D), by recording a two dimensional proton experiment reduced to its diagonal decoupled in the F1 but not F2 dimension and by comparing the intensity of the central line (corresponding to hydrogen atoms linked to ^{14}N) to that of the two satellites (corresponding to hydrogen atoms linked to ^{15}N). By replacing the unlabeled growing medium by the expression labeled medium just prior to the induction, we obtained values of 66% for the DIII-HaV and DIII-RVFV expressed in S2 cells and 63% for actin expressed in Sf9 cells. An explanation for this incomplete labeling could be the presence in the cell, at the moment of their transfer in the labeling medium, of a pool of unlabeled amino acids or amino acid precursors. To test this hypothesis, we introduced a step to purge cells of non-labeled amino acids between the amplification and the protein expression stages by transferring the cells into labeling medium (^{15}N-YE_{10}-ESF) for 48 h and transferring them again in fresh labeling medium just prior to induction. This resulted in a clear increase of the ^{15}N incorporation that reached 77% and 80% in the two assays we performed.

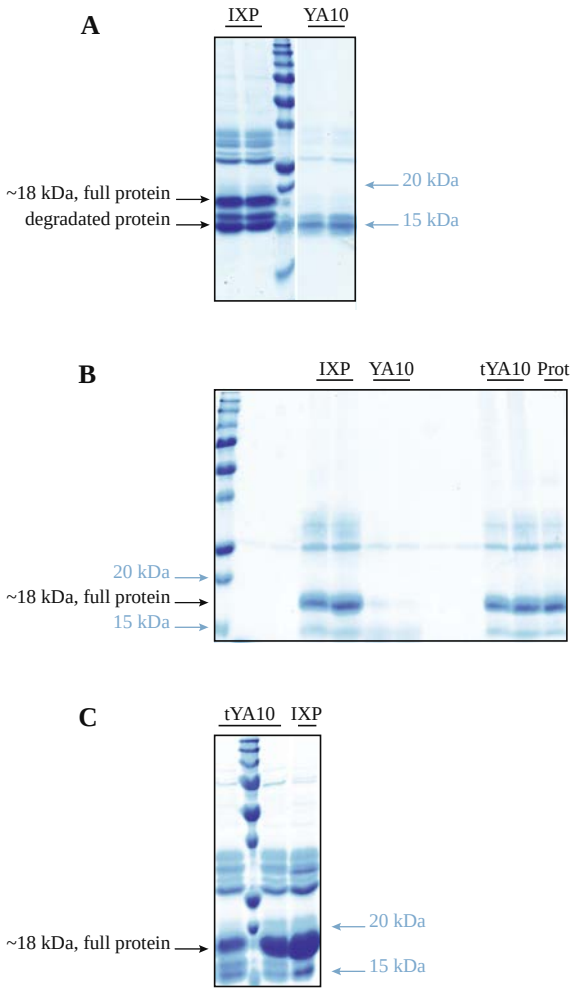


Fig.2. Comparison of the effect of the Insect-Xpress, YA_{10}-ESF and YA_{10}-ESF treated for protease inactivation (tYA_{10}-ESF), on the integrity of a "fragile" protein (DIII-HaV, that was shown to possess an N-terminal unstructured extension, see text). (A) SDS–PAGE analysis of the protein expressed in Insect-Xpress (IXP) and YA_{10}-ESF (YA10). In Insect-Xpress, the major band at 18 kDa corresponds to the full protein and the two others to partially degraded forms (at the level of the N-terminal extremity, see text). Only the most degraded form of the protein is found in the case of the protein expressed in YA_{10}-ESF. (B) SDS–PAGE analysis of the protein expressed in Insect-Xpress, purified and incubated in the three culture media during 48 hours. There is no evolution after incubation in the Insect-Xpress and tYA_{10}-ESF. On the opposite, the protein is degraded after incubation in the untreated YA_{10}-ESF. (C) SDS–PAGE analysis of the protein expressed in Insect-Xpress and tYA_{10}-ESF. The full-length protein is mostly found in both cases.

3.7. NMR

Fig. 3 exemplifies the results obtained by NMR on some systems we tested. We recorded ^{15}N-HSQC spectra for DIII-RVFV, (Fig. 3A) and DIII-HaV (Fig. 3B), and ^{15}N-sofast-HMQC spectra for the non-polymerizable actin mutant (Fig. 3C). All were acquired with protein concentrations of about 100 μM, corresponding to production volumes ranging from 30 ml for DIII-HaV and DIII-RVFV to 250 ml for AP-actin (Table 2). In the case of the DIII-HaV fragments (Fig. 3B), we superimposed the spectra of the degraded (orange, recorded on a sample produced using the non-treated yeast autolysate) and non-degraded (blue, recorded on a sample produced with treated autolysate) proteins, showing that an unstructured fragment of at least 19 amino acids was removed of the degraded protein. This indicates that, provided that the cells are in good condition, the medium reconstituted with treated yeast autolysate is compatible with the production of unstructured proteins, an

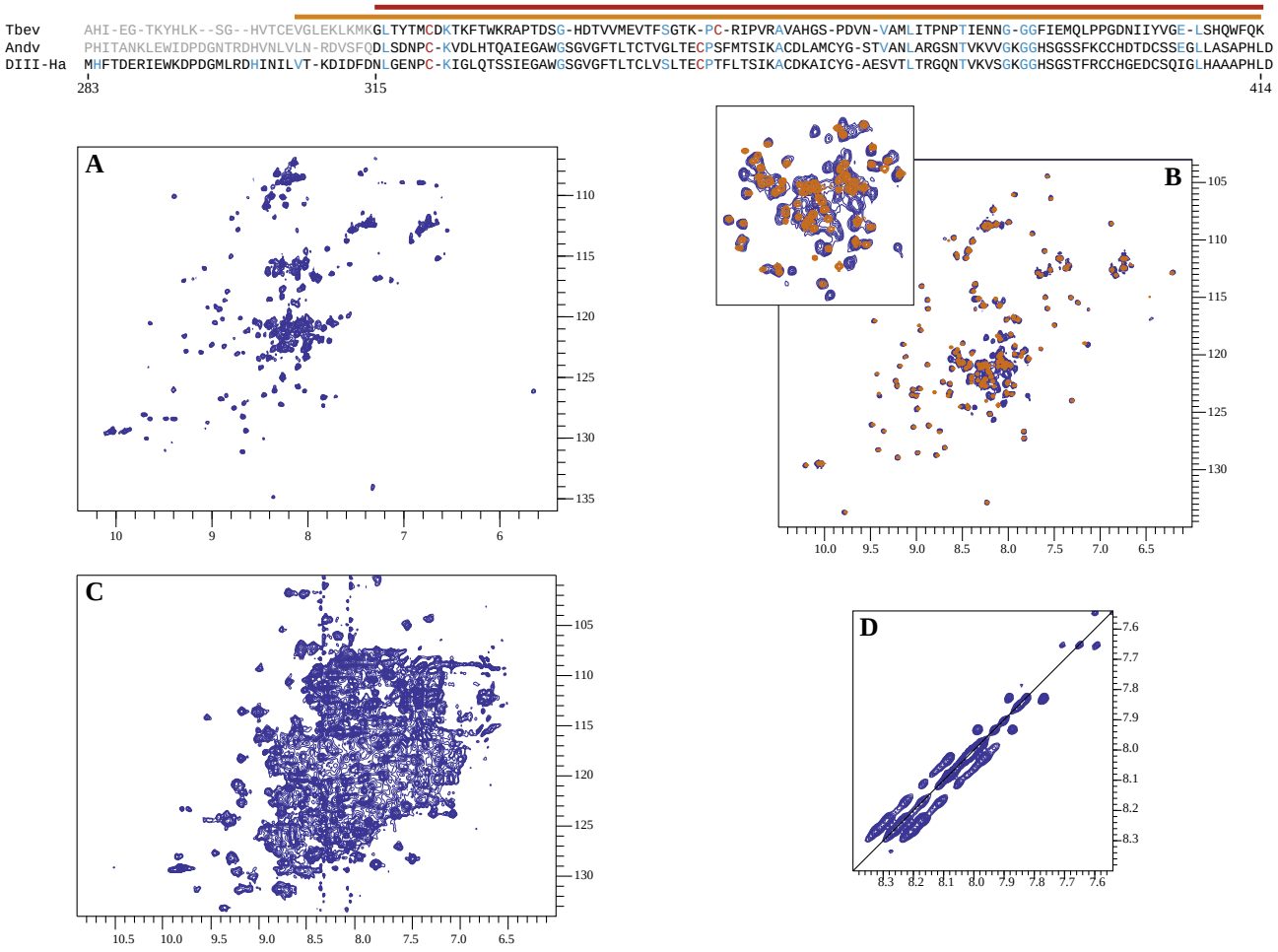


Fig.3. NMR spectra recorded on different proteins produced in YA_{10}-ESF treated for protease inactivation (tYA_{10}-ESF). On the top is the alignment of the sequence of the domain III of the Tick-borne encephalitis virus E protein (the location of the domain III being indicated by the horizontal red line) and of the sequence of the corresponding regions (as predicted by Tischler et al., 2005) of the Andes virus Gc protein and of the DIII-HaV fragment used in this study. The region corresponding to the degraded form of the DIII-HaV fragment (produced with the non-treated yeast autolysate), as delimited by mass spectrometry, is indicated by the orange horizontal line. In (A) is represented the spectrum of DIII-RVFV In (B) are superimposed the HSQC spectra of DIII-HaV produced with untreated YA_{10}-ESF (orange) and treated tYA_{10}-ESF (blue). As shown by the insert, the main differences between the two spectra is the presence of blue additional peaks in the central region, indicating that the difference between the full length and the degraded proteins is the removal of an unstructured fragment at the N-terminal extremity of DIII-HaV. In (C) is represented the spectrum of a non-polymerizable double mutant of actin. In (D) is the correlation spectra used to evaluate the ^{15}N incorporation ratio. The experience in decoupled in F1 but not in F2 ending with three peaks along the F2 (horizontal) dimension for each F1 frequency. The central peak of each triplets corresponds to protons linked to ^{14}N while the two satellite peaks correspond to protons linked to ^{15}N.

important issue for NMR studies. In the case of actin, the size of the protein (42 kDa) precludes the identification of the individual resonances but for some isolated peripheral peaks. The spectrum nevertheless suggests that it will be possible to observe changes induced by regulatory peptide binding upon deuteration (work currently in progress).

With respect to the biological systems, we selected as test case the membrane fusion proteins AFF-1 and EFF-1 from *C. elegans* as well as the Gc from the Rift valley fever virus because recent results have shown that, in spite of lacking amino acids similarity, they have a common fold (the "class II" fold) and derive from a common ancestor (Peres-Varga et al., in press; Dessau and Modis, 2013). Although the ectodomains have between 400 and 600 amino acids in total, depending on the protein, they have an internal domain (domain III or DIII) of about 100 amino acids which displays an immunoglobulin superfamily fold and which plays an important role during the membrane fusogenic process and whose presence is characteristic of the class II proteins. In the case of the Hantaan virus, there is no reported structure available. We only had a prediction by homology modeling made for the Gc protein of a different Hantavirus (Andes virus) (Tischler et al., 2005) based on a sequence alignment with the tick-borne encephalitis flavivirus E protein for which the crystal structure is known (Rey et al., 1995). From this prediction, the authors concluded that Hantavirus Gc proteins are also class II fusion proteins and delineated the domain III from Asp315 to Leu414 (Andes sequence), corresponding to the Asn315-Leu414 fragment of the Hantaan virus Gc. Considering the very low sequence conservation between the Andes Gc protein and the tick-borne encephalitis virus E protein (less than 20% amino acid identity), we tested different constructs by extending the chain at either side of the proposed limits, among which the segment Met283-Asp415 analyzed here.

The spectra of the complete (obtained from treated yeast autolysate) and degraded (from non-treated yeast autolysate) fragments indicate that in both cases, the same structured domain is present, with a large spectral dispersion that is fully compatible with the β-barrel structure expected for an immunoglobulin superfamily domain III of a class II membrane fusion protein. The presence of the tag at the C-terminal end of the protein (clearly identifiable by comparison with the spectrum of DIII-RVFV, produced with

the same tag) strongly indicates that the proteolyzed unstructured amino acids in the degraded form are located at the N-terminus of the fragment, again in agreement with the prediction from the model structure. The N-terminal limit of the degraded form was more precisely located by mass spectrometry as Val307, the predicted beginning of the Hantaan virus Gc domain III being Asn315. The number of additional peaks visible on the spectrum of the complete form (19 peaks) with respect to the degraded form is smaller than the difference in amino acids between the two forms (25 amino acids), but these peaks are located in a very crowded region of the spectrum and some of those missing are likely superimposed with others.

4. Conclusion

We show in this study that it is possible to produce ^{15}N-labeled protein samples in both S2 and Sf9 cells by using an easy, inexpensive and robust method. The medium is obtained by mixing commercial ESF921 depleted in all amino acids, 10 g l^{-1} of (now commercial) labeled yeast extract and 0.3 g l^{-1} (5 mM) of $^{15}NH_4Cl$. The quantity of labeling medium and the practicability of the method are optimized by dissociating the cell growth stage, performed in a standard medium, and the labeling stage, done with a minimal volume of labeling medium without cell adaptation, for both Sf9 and S2 cells. The labeled protein yields obtained under these conditions are reproducibly as good as, or even better than, those obtained in spinners with the commercial medium. The isotopic incorporation is about 65% but can be easily increased up to 80% by purging the cells of non-labeled amino acids after growth in unlabeled medium and before the induction of protein production.

For cost reasons, we were not able to directly compare the performances of our medium with those of the commercial ones. In their first article on Abl kinase labeling using CIL BioExpress 2000 medium, Jahnke and collaborators mention expression yields similar to those obtained in unlabeled commercial media and ^{15}N isotopic incorporation ratios of 90% and 93%. Accordingly, with an isotopic incorporation ratio between 63% and 66% and a similar production yields, it will be necessary to use 1.5 l of our medium to obtain the quantity of ^{15}N-labeled protein produced in one liter of the CIL medium. With a maximal price of 800 € per liter (*Cortecnet* personal communication), our solution remains more than four times cheaper than that of CIL (5000 € for one liter of ^{15}N-BioExpress 2000). This could be further improved by an increase of the incorporation ratio. Our experiments demonstrated that it is possible. Obviously, the way we purged the cells of their ^{14}N-amino acids or amino acid precursors is not economically interesting. However, in a preliminary experiment we observed that the cells tolerate without viability loss an overnight starvation in a medium completely depleted in amino acid before being transferred back into a regular medium. This could provide an inexpensive way to purge the cell of ^{14}N-amino acids before inducing expression in the ^{15}N-labeled medium. Similarly, since the protocol we set up gave satisfactory results for the protein we expressed, we did not explore all the parameters. In particular, it is likely that, varying the cell density at the moment of the transfer into the expression medium or the time length of protein production, it could be possible to further improve the protein yield and/or the incorporation ratio. But it is also likely that this would depend on the particular protein expressed and on the production conditions. Finally, we only used Insect-Xpress as growing medium and amino acid-depleted ESF921 as a basis for our labeling medium, but it is likely that other media would also work.

^{15}N-labeling will not be sufficient for all NMR applications but will nevertheless help in many situations. An HSQC spectrum contains a lot of information on the structural state of a protein. With limited effort and cost, the production of ^{15}N-labeled samples could provide quality control in many situations, allowing, for example, to verify the structural characteristics of a given protein, protein mutant or protein fragment as preamble for a biological or structural study. Here, we have been able to verify that a fragment of the Hantaan Gc protein, designed from a weak alignment, nevertheless contained the expected structured domain, and in conjunction with mass spectrometry, to delineate the limits of the domain, strengthening the hypothesis that Hantavirus Gc protein belongs to the structural class II of fusion proteins. In addition, proteins of up to 100 amino acids have been assigned by using only proton and ^{15}N-edited experiments in the past. Thanks to the development of high field NMR spectrometers, improved analysis software and the possibility to complement the experiments recorded on uniformly ^{15}N-labeled samples by others recorded on specifically ^{15}N or ^{15}N,^{13}C-labeled samples, it is reasonable to envisage the assignment (and maybe the structure determination) of even larger proteins, paving the way to the characterization of many important aspects (dynamics, low affinity interaction, low populated states) of proteins in solution. In addition, we are currently developing a similar approach to allow the production of ^{15}N, ^{2}H-labeled proteins.

Acknowledgments

We thank Jérôme Boibouvier, Pierre Gans (Institut de Biologie Structurale, Grenoble, France), Rudy Ménin (*Bio Springer,* Maisons-Alfort, France) and Nicolas Wolff (Institut Pasteur, Paris, France) for constant fruitful discussions. We also thank Nadine Assrir, Kyu-Ho Park and Laurence Damier-Piolle for numerous advices. We acknowledge the French IR-RMN infrastructure (CNRS FR3050) for access to the Gif-sur-Yvette 950 MHz spectrometer. This work was supported by the French ANRS, the French infrastructure for integrated structural biology (*FRISBI*) and by the European *Instruct* program.

Appendix. Labeling medium preparation (for 300 ml)

All steps, excepted the yeast autolysate powder weighting, are performed in a laminar flow hood.

3 grams of ^{15}N-labeled yeast autolysate powder are weighed in a 500 ml sterile glass bottle.

300 ml of medium (ESF921 depleted in all amino acid) are added under sterile conditions.

0.33 ml of a 3 M $^{15}NH_4Cl$ solution are added under sterile conditions.

The powder is dissolved by gentle manual agitation.

The medium is filtered by using a 0.22 μm Stericup filtration unit and stored in the receiver flask closed under sterile conditions.

The medium is stored at 4 °C.

Appendix. Protein expression in S2 cells (45 ml culture)

All steps, excepted centrifugation and cell incubation, are performed in a laminar flow hood.

140 ml of S2 cells grown in Insect Xpress at a density of $\sim 10^7$ cells ml^{-1} are transferred into a 250 ml sterile disposable centrifuge tube under sterile conditions.

The cells are centrifuged during 5 min at 100 *g*.

The supernatant is discarded and the cells are resuspended in 45 ml of labeling medium under sterile conditions.

18 μl of a 10 mM $CdCl_2$ (4 μM final) solutions are added under sterile conditions.

The 45 ml of cell culture are distributed in three T150 tissue culture flasks.

The flasks are incubated for 7 days at 28 °C.

The cells are pelleted by centrifugation (5 min at 100 *g*) and the cells (case of intracellular proteins) or the supernatant (case of excreted proteins) are collected for purification.

Appendix. Protein expression in Sf9 cells (300 ml culture)

300 ml of Sf9 cells grown in Insect Xpress at a density of $\sim 1.5 \cdot 10^6$ cells ml^{-1} are transferred into two 250 ml sterile disposable centrifuge tube under sterile conditions (laminar flow hood).

The cells are centrifuge during 5 min at 100 *g*.

The supernatant is discarded and the cells are resuspended in the amount baculovirus suspension necessary to obtain a MOI of five under sterile conditions (laminar flow hood).

The two centrifuge tubes are gently shaken during an hour at room temperature by using an horizontal rotative shaker.

The cells are transferred into a 500 ml spinner flask and 300 ml of the labeling medium are added under sterile conditions (laminar flow hood).

The culture is incubated at 28 °C with shaking for 3 days.

The cells are pelleted by centrifugation (5 min at 100 *g*) and the cells (case of intracellular proteins) or the supernatant (case of excreted proteins) are collected for purification.

References

Dessau, M., Modis, Y., 2013. Crystal structure of glycoprotein C from Rift valley fever virus. Proc. Natl. Acad. Sci. U.S.A. 110, 1696–1701.

Egorova-Zachernyuk, T.A., Bosman, G.J.C.G., DeGrip, W.J., 2011. Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. Appl. Microbiol. Biotechnol. 89, 397–406.

Egorova-Zachernyuk, T.A., Bosman, G.J.C.G., Pistorius, A.M.A., DeGrip, W.J., 2009. Production of yeastolates for uniform stable isotope labeling in eukaryotic cell culture. Appl. Microbiol. Biotechnol. 84, 575–581.

Gossert, A.D., Jahnke, W., 2012. Isotope labeling in insect cells. In: Atreya, H.S. (Ed.), Isotope Labeling in Biomolecular NMR. Adv. Exp. Med. Biol. 992, 179–196.

Hamatsu, J., O'Donovan, D., Tanaka, T., Shirai, T., Hourai, Y., Mikawa, T., Ikeya, T., Mishima, M., Boucher, W., Smith, B.O., Laue, E.D., Shirawaka, M., Ito, Y., 2013. High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. J. Am. Chem. Soc. 135, 1688–1691.

Hansen, A.P., Petros, A.M., Mazar, A.P., Pederson, T.M., Rueter, A., Fesik, S.W., 1992. A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. Biochemistry 31, 12713–12718.

Nygaard, R., Zou, Y., Dror, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A., Pan, A.C., Liu, C.W., Fung, J.J., Bokoch, M.P., Thian, F.S., Kobilka, T.S., Shaw, D.E., Mueller, L., Prosser, R.S., Kobilka, B.K., 2013. The dynamic process of β2-adrenergic receptor activation. Cell 152, 532–542.

Öhman, L., Ljunggren, J., Häggström, L., 1995. Induction of a metabolic switch in insect cells by substrate-limited fed batch cultures. Appl. Microbiol. Biotechnol. 43, 1006–1013.

Peres-Varga, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Raveh-Barak, H., Podbilewicz, B., Rey, F.A., 2014. Cell 157, 407–419.

Podbilewicz, B., Leikina, E., Sapir, A., Valansi, C., Suissa, M., Shemer, G., Chernomordik, L.V., 2006. The C. elegans developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and *in vivo*. Dev. Cell 11, 471–481.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. Nature 375, 291–298.

Rould, M.A., Wan, Q., Joel, P.B., Lowey, S., Trybus, K.M., 2006. Crystal structures of expressed non-polymerizable actin in the ADP and ATP states. J. Biol. Chem. 281, 31909–31919.

Sapir, A., Choi, J., Leikina, E., Avinoam, O., Valansi, C., Chernomordik, L.V., Newman, A.P., Podbilewicz, B., 2007. AFF-1, a FOS-1 regulated fusogen, mediates fusion of the anchor cell in *C. elegans*. Dev. Cell 12, 683–698.

Sastry, M., Xu, L., Georgiev, I.S., Bewley, C.A., Nabel, G., Kwong, P.D., 2011. Mammalian production of an isotopically enriched domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy. J. Biomol. NMR 50, 197–207.

Saxena, K., Dutta, A., Klein-Seetharaman, J., Schwalbe, H., 2012. Isotope labeling in insect cells. Methods Mol. Biol. 831, 37–54.

Strauss, A., Bitsch, F., Fendrich, G., Graff, P., Knecht, R., Meyhack, B., Jahnke, W., 2005. Efficient uniform isotope labeling of Abl kinase expressed in baculovirus-infected insect cells. J. Biol. NMR 31, 343–349.

Takashi, H., Shimada, I., 2010. Production of isotopically labeled heterologous proteins in non-*E. coli* prokaryotic and eukaryotic cells. J. Biomol. NMR 46, 3–10.

Tischler, N.D., Gonzalez, A., Perez-Acle, T., Rosemblatt, M., Valenzuela, P.D.T., 2005. Hantaan Gc glycoprotein: evidence for a class II fusion protein. J. Gen. Virol. 86, 2937–2947.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPNMR data model for NMR spectroscopy: development of a software pipeline. Proteins 59, 687–696.

Walton, W., Kesprazak, A.J., Hare, J.T., Logan, T.M., 2006. An economic approach to isotopic enrichment of glycoproteins expressed from Sf9 insect cells. J. Biol. NMR 36, 225–233.

Werner, K., Richter, C., Klein-Seetharaman, J., Schwalbe, H., 2008. Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. J. Biomol. NMR 40, 49–53.

Titre Etude des états multiples des domaines WH2 en interaction avec l'actine par résonance magnétique nucléaire

Résumé Les domaines thymosineβ/WH2 sont une famille de protéines intrinsèquement désordonnées impliqués dans le remodelage du cytosquelette d'actine. Ces domaines de 20 à 50 acides aminés existent seuls ou au sein de protéines modulaires, isolés ou répétés. Ils exercent de nombreuses fonctions : ils séquestrent des monomères d'actine, promeuvent l'assemblage unidirectionnel du filament, nucléent, fragmentent et coiffent les filaments. Tous les domaines WH2 interagissent de manière similaire avec l'actine via une hélice amphipathique N-terminale suivie d'un brin central et d'une région C-terminal plus ou moins longue et dynamique. Une étude antérieure a montré que la fonction des domaines βT/WH2 isolés était liée à la dynamique du complexe avec l'actine déterminée par une combinaison d'interactions intermoléculaires le long de l'ensemble de la séquence. Mais les mécanismes expliquant la multifonctionnalité des domaines WH2 répétés restent vagues. Ce travail de thèse présente tout d'abord la production d'actine recombinante, sauvage et mutée dans le système baculovirus/Sf9 pour la biologie structurale ainsi que le développement de stratégies de marquage isotopique en cellules d'insectes. La deuxième partie s'intéresse à la caractérisation structurale et dynamique de domaines WH2 seules en solution : deux domaines isolés et deux protéines contenant deux domaines WH2. Les hélices amphipathiques N-terminales sont partiellement repliées avec des populations variant selon les protéines. La préstructuration des régions C-terminales est plus variable, complètement désordonnée ou partiellement hélicoïdale selon les protéines. La dernière partie présente l'étude de l'interaction de ces protéines avec l'actine.

Mots-clés actine, domaines WH2, N-WASP, Corbon-bleu, protéines intrinsèquement désordonnées, résonance magnétique nucléaire

Title Interaction mechanisms of intrinsically disordered WH2 repeats with actin by nuclear magnetic resonance spectroscopy

Abstract WH2 domains are a family of intrinsically disordered proteins involved in actin cytoskeleton remodeling. These short domains, isolated or repeated in various actin binding proteins display a low sequence identity and a large panel of functions such as sequestration of actin monomers, promotion of unidirectional assembly, nucleation, fragmentation, filament capping. All WH2 domains fold similarly upon actin binding. They form an extended interface along actin, with an amphipathic N-terminal helix followed by an extended central strand and a more dynamic C-terminal region. Previous work on single βT/WH2 domains showed that function was linked to the dynamics of the complex with actin which is determined by a combination of intermolecular interactions throughout the sequence. The multifunctionality of WH2 tandem repeats is still elusive. The present work first describes production of recombinant wild-type and mutant actin in insect cells and isotopic ^{15}N-labeling for NMR spectroscopy. As a first step to gain insight into the folding upon binding mechanism of functionally different WH2 repeats, we investigated the conformational behavior of two single domains and two tandem repeats free in solution by NMR. The N-terminal amphipatic helix is partially formed but with various propensities depending on the proteins while the C-terminal region that may form an helix in the complex may be either completely disordered or partially formed in absence of actin. Investigation of WH2:actin interaction for the same four proteins is described in the last chapter.

Keywords actin, WH2 repeats, N-WASP, Corbon-bleu, intrinsically disordered proteins, nuclear magnetic resonance